FeedNetBack - D03.02 - Control Subject to Transmission Constraints, With Transmission Errors

Federica Garin, Sandro Zampieri, Lei Bao, Carlos Canudas de Wit, Ruggero Carli, Tobias Oechtering, Mikael Skoglund, Ali Abbas Zaidi

GRANT AGREEMENT N°223866

Deliverable	Deliverable D03.02
Nature	Deliverable
Dissemination	Public

D03.02 - CONTROL SUBJECT TO TRANSMISSION CONSTRAINTS, WITH TRANSMISSION ERRORS

Report Preparation Date	04/08/2011
	Project month: 36
Authors	Federica Garin, INRIA Grenoble
	Sandro Zampieri, Università di Padova
	Lei Bao, KTH
	Carlos Canudas-de-Wit, INRIA Grenoble
	Ruggero Carli, Università di Padova
	Tobias J. Oechtering, KTH
	Mikael Skoglund, KTH
	Ali A. Zaidi, KTH
Report Version	Vn 1
Doc ID Code	INRIA-UNIPD_D03.02_V01
Contract Start Date	01/SEP/2008
Duration	41 months
Project Coordinator :	Carlos CANUDAS DE WIT, INRIA, France

SEVENTH FRAMEWORK PROGRAMME

Theme 3:

Information and Communication Technologies

SUMMARY

This Deliverable Report describes the research performed within Work Package 3, Task 3.2 (Control Subject to Transmission Constraints, with Transmission Errors), in the first 36 months of the project. It targets the issue of control subject to transmission constraints with transmission error. This research concerns problems arising from the presence of a noisy communication channel (specified and modeled at the physical layer) within the control loop. The resulting constraints include finite capacities in the transmission of the sensor and/or actuator signals and transmission errors. Our focus is on designing new compression and coding techniques to support networked control in this scenario.

This Deliverable extends the analysis provided in the companion Deliverable D03.01, to deal with the effects of noise in communication channel. The quantization schemes described in D03.01, in particular the adaptive ones, might be very sensitive to the presence of even a few errors. Indeed error-correction coding for estimation or control purposes cannot simply exploit classical coding theory and practice, where vanishing error probability is obtained only in the limit of infinite block-length.

A first contribution reported in this Deliverable is the construction of families of codes having the any-time property required in this setting, and the analysis of the trade-off between code complexity and performance. Our results consider the binary erasure channel, and can be extended to more general binary-input output-symmetric memoryless channels.

The second and third contributions reported in this deliverable deal with the problem of remotely stabilizing linear time invariant (LTI) systems over Gaussian channels. Specifically, in the second contribution we consider a single LTI system which has to be stabilized by a remote controller using a network of sensors having average transmit power constraints. We study basic sensor network topologies and provide necessary and sufficient conditions for mean square stabilization.

Then in the third contribution, we extend our study to two LTI systems which are to be simultaneously stabilized. In this regard, we study the interesting setups of joint and separate sensing and control. By joint sensing we mean that there exists a common sensor node to simultaneously transmit the sensed state processes of the two plants and by joint control we mean that there is a common controller for both plants. We name these setups as: i) control over multiple-access channel (separate sensors, joint controller setup), ii) control over broadcast channel (common sensor, separate controllers setup), and iii) control over interference channel (separate sensors, separate controllers). We propose to use delay-free linear

schemes for these setups and thus obtain sufficient conditions for mean square stabilization.

Then, we discuss the joint design of the encoder and the controller. We propose an iterative design procedure for a joint design of the sensor measurement quantization, channel error protection, and controller actuation, with the objective to minimize the expected linear quadratic cost over a finite horizon.

Finally, the same as for the noiseless case, we address the issues that arise when not only one plant and one controller are communicating through a channel, but there is a whole network of sensors and actuators. We consider the effects of digital noisy channels on the consensus algorithm, and we present an algorithm which exploits the any-time codes discussed above.

Contents

1 Introduction

This Deliverable targets the issue of control subject to transmission constraints with transmission error. It is worth recalling that the case with no transmission error is analyzed in the companion Deliverable D03.01, where the assumption of noiseless channel is essential. Indeed, most results illustrated in D03.01, and in particular the zoom-in zoom-out coding strategy, require a perfect synchronization of the encoder's and decoder's internal state, which is lost if even one single bit is wrongly received. However, in communication settings such as wireless sensor networks, the assumption of a noiseless communication becomes irrealistic. It would be tempting to separate the error-correcting problem from the (quantized) control problem, but classical results in coding theory allow to achieve vanishing error probability only in the limit of infinitely large codeword length, which is impossible to use within an estimation or control task. A series of works by A. Sahai and S. Mitter [115–117], show the specific feature distinguishing the problem of information transmission for control from the problem of pure information transmission, related to the different sensitivity to delay. Indeed, while the presence of sensible delays can often be tolerated in the communication performance evaluation, such delays can be detrimental for the system performance in several control applications. Here, the fundamental question is not only where, but also when the information is available. For this reason, it is often desirable to use transmission systems for control applications which are able to provide estimates whose precision increases with time, so as providing a reasonable partial information transmission anytime the process is stopped. In section 2, we present families of coding schemes having this anytime property, and we discuss the tradeoff between their performance and their complexity. This is the first result illustrated in this report.

In section 3 we study the problem of remotely stabilizing a first order LTI system over Gaussian sensor networks. It is worth remarking that the the problem of stabilization of a LTI system, is the most studied problem in the literature about control over noisy channels (see e.g. [60, 102, 116]). One original aspect of our approach is in the channel model considered, i.e., the relay channel, which is more realistic for wireless networks. In our setup we assume that there exists a network of sensor nodes which communicate the state process to a remote control unit, which takes actions to stabilize the plant. All the communication links between the plant, the sensors, and the controller are assumed to be corrupted by additive white Gaussian noise. We study some basic network topologies for the sensors and derive necessary and sufficient conditions for mean square stabilization over these basic sensor network settings. Our results reveal relationship between mean square stability of the plant and the communication channel parameters, e.g., sensor power, noise power, channel gains etc.

Then in section 4 we study the problem of simultaneously stabilizing two first order LTI systems over three basic multi-user communication channels: i) white Gaussian multiple-access channel, ii) white Gaussian broadcast channel, and iii) white Gaussian interference channel. We propose to use linear and memoryless communication and control schemes over these channels and thereby derive sufficient conditions for mean square stabilization. These delay-free linear schemes are inspired by the well-known Schakwijk-Kailath coding scheme [118], and they are suitable for delay-sensitive control applications.

As previously mentioned, most work on control over noisy channels has been devoted to stability, while optimal designs are much less explored in the literature. Exceptions include the study of optimal stochastic control over communication channels, e.g., [68, 83, 103, 124]. In section 5, our main concern is optimal average performance over a finite horizon, given a fixed data rate. We introduced an iterative design procedure for finding encoder–controller pairs. The result is a synthesis technique for joint optimization of the quantization, error protection and control over a bandlimited and noisy channel.

Finally, the same as for the noiseless case, we address the issues that arise when not only one plant and one controller are communicating through a channel, but there is a whole network of sensors and actuators. In section 6, we consider the effects of digital noisy channels on the consensus algorithm, and we present an algorithm which exploits the any-time codes presented in section 2.

2 Anytime coding algorithms for estimation through a noisy digital channel

In this section we will report the work presented in [70] on anytime coding algorithms for estimation through a noisy digital channel, which are a fundamental building block of any estimation or control algorithm over a network affected by noise on the transmission channels. Reliable transmission of information among the nodes of a network is known to be a relevant problem in information engineering. It is indeed fundamental both when the network is designed for pure information transmission, as well as in scenarios in which the network is deputed to accomplish some specific tasks requiring information exchange, such as parallel and distributed computation, or load balancing; wireless sensor networks; sensors/actuators networks, such as mobile multi-agent networks. Distributed algorithms to accomplish synchronization, estimation, or localization tasks, necessarily need to exchange quantities among the agents, which are often real-valued. Assuming that transmission links are digital, a fundamental problem is thus to transmit a continuous quantity, i.e. a real number or, possibly, a vector, through a digital noisy channel up to a given degree of precision.

In [70] we addressed the problem of efficient, real-time transmission of a finite-dimensional Euclidean-space-valued state through a noisy digital channel. We focused on anytime transmission algorithms, i.e. algorithms which can be stopped anytime while providing estimations of increasing precision. These algorithms are particularly suitable for applications in problems of distributed control.

As especially pointed out in a series of works by A. Sahai and S. Mitter [115–117], there is a specific feature distinguishing the problem of information transmission for control from the problem of pure information transmission. This is related to the different sensitivity to delay typically occurring in the two scenarios. Indeed, while the presence of sensible delays can often be tolerated in the communication performance evaluation, such delays can be detrimental for the system performance in several control applications. Here, the fundamental question is not only where, but also when the information is available. For this reason, it is often desirable to use transmission systems for control applications which are able to provide estimates whose precision increases with time, so as providing a reasonable partial information transmission anytime the process is stopped.

On the other hand, the computational complexity of the transmission schemes is a central issue. In fact, nodes in wireless networks are usually very simple devices with limited computational abilities and severe energy constraints. Applicable transmission systems should be designed performing a number of operations which remains bounded in time, both in the encoding and in the decoding. Hence, an analysis of the tradeoffs between performance and complexity of the transmission schemes is required.

In many problems of information transmission, there is the possibility to take advantage of the feedback information naturally available to the transmitter. Known results in Information Theory [71] show that feedback can improve the capacity of channels with memory, or multiple access channels[1], as well as reduce latency and computational complexity. In many cases of practical interest, however, feedback information is incomplete, or difficult to be used. Also, there are many situations, for instance in the wireless network scenario, in which the transmitter needs to broadcast its information to many different receivers and hence feedback strategies to acknowledge the receipt of past transmissions could be unfeasible. For these reasons, in the present work we shall restrict ourselves to the case in which there is no feedback information.

A fundamental characteristic of digital communication for control applications concerns the nature of information bits. In the traditional communication theory, information bits are usually assumed to be equally valuable, and they are consequently given the same priority by the transmission system designer. In fact, design paradigms of modern low-complexity codes [104, 114] –based on random sparse graphical models and iterative decoding algorithms– treat information bits as equally valuable. While such an assumption is typically justified by the source-channel separation principle, this principle does not generally hold when delay is a primary concern. For instance, it is known that separate source-channel coding is suboptimal in terms of the joint source-channel error exponent [72, 73]. In fact, in many problems of information transmission for control or estimation, different information bits typically require significantly different treatment.

As an example, particularly relevant for the topics addressed in this section, assume that a random parameter, uniformly distributed over a unitary interval, has to be reliably transmitted through a digital noisy channel (see [63] and references therein for the analysis of the information theoretic limits of this problem on the bandwidth-unlimited Gaussian channel). Such a parameter may be represented by its dyadic expansion, which is a stream of independent identically distributed bits. Clearly, such information bits are not equally valuable, since the first one is more significant than the second one, and so on. This motivates the study of

[1] Whereas a classic result due to Shannon shows that feedback does not improve the capacity of a discrete memoryless channel.

unequal error protection codes [58, 100]. One of the challenges posed by information transmission for control/estimation applications is to come up with design paradigms for practical, low-complexity, unequal error protection codes.

In [70] we proposed two classes of coding strategies for the anytime transmission of real-valued random vectors through a digital noisy channel. In both cases, the transmission scheme consists of an encoder, mapping the real vector into a sequence of channel inputs, and of a decoder, sequentially refining the estimate of the vector as more and more channel outputs are observed. The first strategy is characterized by good performance in terms of the convergence of the mean squared error, but it is expensive in terms of encoder/decoder computational complexity. On the other hand, the second class of strategies have convenient computational complexity, but worse convergence rate.

In order to keep the use of information-theoretical techniques at a minimum, we shall confine our exposition to the binary erasure channel (BEC), and defer any discussion on the possible extensions to general discrete memoryless channels to the concluding section. In the BEC, a transmitted binary signal is either correctly received, or erased with some probability ε. While this channel allows for an elementary treatment, it is of its own interest in many scenarios. In [64], the techniques proposed here have been applied in order to obtain a version of the average consensus algorithm working in presence of digital erasure communication channels between the nodes.

We end this introduction by establishing some notation. Throughout the section, $\mathbb{R}$ and $\mathbb{N}$ will denote the sets of reals and naturals, respectively. For a subset $A \subseteq B$, $|A|$ will denote the cardinality of A, $\overline{A} = B \setminus A$ its complement, and $\mathbf{1}_A : B \to \{0,1\}$ its indicator function, defined by $\mathbf{1}_A(x) = 1$ if $x \in A$, and $\mathbf{1}_A(x) = 0$ otherwise. The natural logarithm will be denoted by $\ln$, while $\log$ will stand for the logarithm in base 2. For $x \in [0,1]$, we shall use the notation $\mathrm{H}(x) := -x \log x - (1-x)\log(1-x)$ for the binary entropy of x with the standard convention $0 \log 0 = 0$. For two sequences of reals $(a_t)_{t\in\mathbb{N}}$ and $(b_t)_{t\in\mathbb{N}}$, both the notations $a_t = O(b_t)$ and $b_t = \Theta(a_t)$ will mean that $a_t \le K b_t$ for some constant K, while $a_t = o(b_t)$ will mean that $\lim_t a_t/b_t = 0$. A sequence a_t, $t = 1, 2, \ldots$ is sometimes denoted with the symbol $a = (a_t)_{t=1}^{\infty}$, while with the symbol $a = (a_t)_{t=1}^{T}$ we will mean its truncation to $t = 1, \ldots, T$.

2.1 Problem formulation

We shall now provide a formal description of the problem. Let x be a random variable taking values on $\mathcal{X} \subseteq \mathbb{R}^d$. We shall assume that x has an a priori probability law which is absolutely continuous with respect to the Lebesgue measure, and denote by $f(x)$ the probability density of x. Further, we shall assume that $\mathbb{E}||x||^{2+\delta} < +\infty$ for some $\delta > 0$. At time $t \in \mathbb{N}$, the communication channel has input y_t, and output z_t, taking values in some finite alphabets $\mathcal{Y}$, and $\mathcal{Z}$, respectively. Transmission is assumed to be memoryless, i.e., given the current input y_t, the output z_t is assumed to be conditionally independent from the previous inputs $(y_s)_{s=1}^{t-1}$ and outputs $(z_s)_{s=1}^{t-1}$, as well as from the vector x. The conditional probability of $z_t = z$ given $y_t = y$ will be assumed stationary and denoted by $p(z|y)$. We shall consider in detail the binary erasure channel (BEC) in which $\mathcal{Y} = \{0, 1\}$, $\mathcal{Z} = \{0, 1, ?\}$, and

$$p(?|0) = p(?|1) = \varepsilon\,, \quad p(0|0) = p(1|1) = 1 - \varepsilon\,, \quad p(1|0) = p(0|1) = 0\,.$$

Here, $?$ stands for the erased signal, and $\varepsilon \in [0, 1[$ for the erasure probability.

The anytime transmission scheme consists of an encoder and a sequential decoder.[2] The encoder consists of a family of maps $E_t : \mathcal{X} \to \mathcal{Y}$, specifying the symbol transmitted through the channel at time t, $y_t = E_t(x)$. With this family of maps we can associate the global map $\mathcal{E} : \mathcal{X} \to \mathcal{Y}^{\mathbb{N}}$ which specifies the infinite string that the encoder generates from x. The decoder instead is given by a family of maps $\mathcal{D}_t : \mathcal{Z}^t \to \mathcal{X}$, describing the estimate $\hat{x}_t = \mathcal{D}_t((z_s)_{s=1}^t)$ of x obtained from the string $(z_s)_{s=1}^t$ that has been received until time t. With this family of maps we can associate naturally the global map $\mathcal{D} : \mathcal{Z}^{\mathbb{N}} \to \mathcal{X}^{\mathbb{N}}$. This is represented in the following scheme

$$\begin{array}{ccccccc} \mathcal{X} & \xrightarrow{\mathcal{E}_t} & \mathcal{Y}^t & \xrightarrow{\text{Channel}} & \mathcal{Z}^t & \xrightarrow{\mathcal{D}_t} & \mathcal{X} \\ \\ x & \longmapsto & (y_s)_{s=1}^t & \longmapsto & (z_s)_{s=1}^t & \longmapsto & \hat{x}_t \end{array} \tag{1}$$

where $\mathcal{E}_t := \pi_t \circ \mathcal{E}$ and where $\pi_t : \mathcal{Y}^{\mathbb{N}} \to \mathcal{Y}^t$ is the projection of a sequence in $\mathcal{Y}^{\mathbb{N}}$ into its first t symbols

In order to evaluate the performance of a scheme, we define the root mean squared error (mean with respect to both the randomness of $x \in \mathcal{X}$ and with respect to the possible randomness of the communication channel) at time t by

$$\Delta_t := (\mathbb{E}||x - \hat{x}_t||^2)^{1/2}\,. \tag{2}$$

[2] Our definition of anytime transmission scheme does not formally coincide with that in the Anytime Information Theory of S. Mitter and A. Sahai. Our usage of the term “anytime” has to be understood in the broader sense it has in Artificial Intelligence, where anytime algorithms are algorithms whose quality of results improves gradually as computation time increases [135].

In this section, we shall be concerned with the rate of decay of Δ_t for different anytime transmission schemes. All the coding strategies which will be analyzed are characterized by a root mean squared error Δ_t converging to zero like $2^{-\beta t^\alpha}$ for some constants $\beta > 0$ and $0 < \alpha \leq 1$. More precisely we shall seek to find α, β such that

$$\Delta_t \leq p(t) 2^{-\beta t^\alpha} \tag{3}$$

for some polynomial $p(t)$. When (3) holds, the coding strategy will be said to achieve a degree of convergence α and rate of convergence β. When $\alpha = 1$ we shall simply say that we have an exponential convergence. In this case β is referred to as the exponential convergence rate. In the sequel, various strategies will be compared in terms of the parameters α and β that can be achieved, and such parameters will be related to the required computational complexity.

2.2 Application to state estimation under communication constraints

The problem illustrated in the previous paragraph is related to the state estimation problem under communication constraints (see [101, 102, 121, 123, 124] and references therein). Assume we are given a discrete time stochastic linear system

$$x(t+1) = Ax(t) + v(t), \qquad x(0) = x_0, \tag{4}$$

where $x_0 \in \mathbb{R}^n$ is a random vector with zero mean, $v(t) \in \mathbb{R}^n$ is a zero-mean white noise, $x(t) \in \mathbb{R}^n$ is the state sequence, and $A \in \mathbb{R}^{n \times n}$ is a full rank, unstable matrix.

Suppose that a remotely positioned receiver is required to estimate the state of the system, while observing the output of a binary erasure channel only. Then, it is necessary to design a family of encoders E_t and of decoders D_t. At each time $t \geq 0$, the encoder E_t takes $x(0), \dots, x(t)$ as input, and returns the symbol $y_t \in \{0, 1\}$, which is in turn fed as an input to the channel. The receiver observes the channel output symbols $z_0, \dots, z_t$, from which the decoder D_t has to obtain an estimate $\hat{x}(t)$ of the current state.

If we have that $v(t) = 0$ for every $t \geq 0$, then the only source of uncertainty is due to the initial condition x_0. Hence, in this case, the encoder/decoder task reduces to obtaining good estimates of x_0 at the receiver side. Indeed, in order to obtain a good estimate $\hat{x}(t)$ of $x(t)$, the receiver has to obtain the best possible estimate $\hat{x}(0|t)$ of the initial condition $x(0)$ from the received data $y_0, \dots, y_t$, and then it can define $\hat{x}(t) := A^t \hat{x}(0|t)$. In this way, one has $x(t) - \hat{x}(t) = A^t(x(0) -$

$\hat{x}(0|t)$), so that the problem reduces to finding the best way of coding $x(0)$ in such a way that expansion of A^t is well dominated by the contraction of $x(0) - \hat{x}(0|t)$. The same technique can be applied if $v(t)$ is small with respect to x_0 as clarified by the following example.

Example 1. *Consider the following unstable scalar discrete time linear system*

$$x(t+1) \;=\; ax(t) + v(t)\,, \qquad x(0) = x_0\,,$$

where $a > 1$ and where x_0 is a random variable with probability density $f(x)$ and $v(t)$ is a sequence of independent, identically distributed random variables with zero mean and variance σ_v^2, which are independent of x_0. Assume that a state estimation algorithm is run, based on the noiseless model $x(t+1) = ax(t)$ by estimating the initial condition x_0 from data transmitted until time t. As before, we shall denote this estimate by $\hat{x}(0|t)$. From $\hat{x}(0|t)$, we form the estimate $\hat{x}(t) := a^t\hat{x}(0|t)$ of $x(t)$. The estimation error at time t will be $e(t) := x(t) - \hat{x}(t) = a^t(x(0) - \hat{x}(0|t)) + \sum_{i=0}^{t-1} a^{t-1-i}v(i)$, so that $\mathbb{E}[e(t)^2] = a^{2t}\mathbb{E}[(x(0) - \hat{x}(0|t))^2] + \sigma_v^2\frac{1-a^{2t}}{1-a^2}$. This error depends both on the error in the estimation of the initial condition, and on the wrong model we used. As we shall see, our techniques yield an estimation error on $x(0)$ of the form $\mathbb{E}[(x(0) - \hat{x}(0|t))^2] = C\zeta(t)$, where C depends only on the probability density $f(x)$ and $\zeta(t)$ is a function converging to zero depending only on the communication channel characteristics and on the coding strategy. Therefore,

$$\mathbb{E}[e(t)^2] = a^{2t}\left[C\zeta(t) + \sigma_v^2\frac{1-a^{-2t}}{a^2-1}\right].$$

In case C is much larger than σ_v^2, there will be an initial time regime in which the error is not influenced by the model noise but only by the estimation of the initial condition $x(0)$.

2.3 The limit of performance on the binary erasure channel

Observe that, in case of noiseless channel, the function mapping x into $(\hat{x}_0, \ldots, \hat{x}_{t-1})$ is a quantizer assuming at most 2^t values. It is well-known in the theory of vector quantization [86] that, if $\mathcal{Q} : \mathcal{X} \to \mathcal{X}$ is a quantizer assuming m values, then

$$(\mathbb{E}||x - \mathcal{Q}(x)||^2)^{1/2} \geq C_{-}m^{-1/d}\,, \tag{5}$$

where C_{-} is a positive constant only depending on the dimension d, and the a priori density $f(x)$. This shows that $\Delta_t \geq C_{-}2^{-t/d}$ for all $t \in \mathbb{N}$. Hence, it is not possible to obtain a convergence degree α greater than 1 with an exponential convergence rate β larger than $1/d$. In this section, we shall present a tighter upper bound on the exponential convergence rate of Δ_t on the BEC with erasure probability ε.

Consider the general scheme (1). The error pattern associated to the output sequence $(z_t) \in \mathcal{Z}^{\mathbb{N}}$ is the sequence $(\xi_t) \in \{c, ?\}^{\mathbb{N}}$ componentwise defined by $\xi_t = c$ if $z_t \in \{0, 1\}$ (this corresponds to a correct transmission), and $\xi_t = ?$ if $z_t = ?$ (this corresponds to an erased signal). Observe that, given the encoder $\mathcal{E}$ and the decoder $\mathcal{D}$, the error pattern $(\xi_t)_{t\in\mathbb{N}}$ is a random variable independent of the source vector x. This property will allow us to present for the BEC almost elementary proofs of results holding true also for general discrete memoryless channels.

For $j \leq t$, let

$$\lambda_j^t := \textstyle\sum_{j\leq s\leq t} \mathbf{1}_{\{\xi_s = c\}} \tag{6}$$

be the random variable describing the number of non-erased outputs observed between time j and t. Clearly,

$$\mathbb{P}(\lambda_j^t = l) = \binom{t-j+1}{l} \varepsilon^{t-j+1-l}(1-\varepsilon)^l \,, \qquad l = 0, \dots, t-j+1 \,. \tag{7}$$

The simple observation above allows one to prove the following result.

Theorem 1. *Assume transmission over the BEC with erasure probability $\varepsilon \in [0, 1]$. Then, the estimation error of any coding scheme as in (1) satisfies*

$$\Delta_t \geq C_- \, 2^{-t\overline{\beta}(d,\varepsilon)} \,, \tag{8}$$

for all $t \geq 0$, where

$$\overline{\beta}(d, \varepsilon) := -\frac{1}{2} \log \left(\varepsilon + (1-\varepsilon)2^{-2/d}\right) \tag{9}$$

and C_- is a constant depending only on the dimension d and the a priori density $f(x)$.

It is not hard to see that (8) continues to hold true even if the encoder has access to noiseless (even non-causal) output feedback.[3] A fortiori, (8) holds in the case of

[3] In fact, it is tempting to conjecture that a tighter bound could possibly be proven for the exponent in the absence of feedback.

partial or noisy feedback, which is the typical situation occurring in the network scenarios outlined above. In case of perfect causal feedback, the bound (8) is achieved by the encoder which keeps on transmitting the most significant bit of the dyadic expansion of x until this is correctly received. However, it is not clear what can be done if the feedback is noisy, partial, or not available. We shall propose some simple schemes which are not able to achieve exponential error rates, but have low computational complexity, while in the sequel we shall present schemes achieving exponential error rates at the cost of higher computational complexity.

2.4 Quantized encoding schemes

In this section, we shall propose and compare different coding strategies. All of them are based on a separation between the quantization of the continuous vector and the channel coding. In the literature, vector quantizers with special structure have been proposed, called tree-structured vector quantizers [86]. Consider a map $\mathcal{S} : \mathcal{X} \to \{0,1\}^{\mathbb{N}}$, and, for all $t \in \mathbb{N}$, the map $\mathcal{S}_t := \pi_t \circ \mathcal{S}$, where $\pi_t : \{0,1\}^{\mathbb{N}} \to \{0,1\}^t$ is the truncation operator defined above. Finally, let $\mathcal{S}_t^{-1}$ be a right inverse of $\mathcal{S}_t$. Then, we can define a tree-structured vector quantizer [86, pag.410] which is the family of maps $\mathcal{Q}_t : \mathcal{X} \to \mathcal{X}$ defined as $\mathcal{Q}_t := \mathcal{S}_t^{-1} \circ \mathcal{S}_t$. It can be seen [86] that if $\mathbb{E}||x||^{2+\delta} < +\infty$ for some $\delta > 0$, then, there exists a tree-structured vector quantizer $(\mathcal{Q}_t)$ such that

$$(\mathbb{E}||x - \mathcal{Q}_t(x)||^2)^{1/2} \leq C_+ 2^{-t/d}\,, \tag{10}$$

where C_+ is a positive constant depending only on the dimension d and the a priori density $f(x)$. Observe that the right-hand side of (10) differs from the right-hand side of (5) only by a constant independent of the quantizer's range size $m = 2^t$, i.e. tree-structured quantizers are not suboptimal for their rate of convergence.

Remark 1. *The upper bound (10) is easy to be obtained if $\mathcal{X} = [0,1]$. Indeed, in this case, one can take $\mathcal{S}$ to be the map which associates with x its binary expansion. We can apply this argument in case $\mathcal{X}$ is a bounded subset of $\mathbb{R}$. In case $\mathcal{X}$ is unbounded, tree-structured quantizers can be determined satisfying the upper bound (10) (see Lemma 5.2 in [109]). The extension from the scalar to the vector case is straightforward.*

Notice that, if $x', x'' \in \mathcal{X}$ are such that $\mathcal{S}_t(x') = \mathcal{S}_t(x'')$, then

$$\begin{aligned} \mathbb{E}||x' - x''||^2 &\leq \mathbb{E}(||x' - \mathcal{Q}_t(x')|| + ||x'' - \mathcal{Q}_t(x'')||)^2 \\ &\leq 2\mathbb{E}||x' - \mathcal{Q}_t(x')||^2 + 2\mathbb{E}||x'' - \mathcal{Q}_t(x'')||^2 \leq 2C_+^2 2^{-2t/d}\,. \end{aligned} \tag{11}$$

With a slight abuse of terminology, the map $\mathcal{S}$ associated with a tree-structured vector quantizer will be called a dyadic expansion map. We now show how a transmission scheme can be built starting from $\mathcal{S}$ and a family of its truncations' right inverses $\mathcal{S}_t^{-1}$.

Consider a sequence of integers $m_1, m_2, \ldots \in \mathbb{N}$ such that $m_{t-1} \leq m_t$ for all t and a family of maps

$$\tilde{E}_t : \mathcal{Y}^{m_t} \to \mathcal{Y}\,, \qquad \tilde{\mathcal{D}}_t : \mathcal{Z}^t \to \mathcal{Y}^{m_t}\,. \tag{12}$$

We can define the map $\tilde{\mathcal{E}} : \mathcal{Y}^{\mathbb{N}} \to \mathcal{Y}^{\mathbb{N}}$ by letting the value of $\tilde{\mathcal{E}}((w_s)_{s=1}^{\infty})$ at time t equal to $\tilde{E}_t(w_1, \ldots, w_{m_t})$. We also put $\tilde{\mathcal{E}}_t := \pi_t \circ \tilde{\mathcal{E}}$. Notice that, since $\tilde{\mathcal{E}}_t((w_s)_{s=1}^{\infty}))$ depends on $w_1, \ldots, w_{m_t}$ only, then $\tilde{\mathcal{E}}_t$ is actually a map from $\mathcal{Y}^{m_t}$ to $\mathcal{Y}^t$. Finally encoders and decoders are defined by $\mathcal{E}_t := \tilde{\mathcal{E}}_t \circ \mathcal{S}_{m_t}$ and $\mathcal{D}_t := \mathcal{S}_{m_t}^{-1} \circ \tilde{\mathcal{D}}_t$. The overall sequence of maps is described by the following scheme

$$\begin{array}{ccccccccccc} \mathcal{X} & \xrightarrow{\mathcal{S}_{m_t}} & \mathcal{Y}^{m_t} & \xrightarrow{\tilde{\mathcal{E}}_t} & \mathcal{Y}^t & \xrightarrow{\text{Channel}} & \mathcal{Z}^t & \xrightarrow{\tilde{\mathcal{D}}_t} & \mathcal{Y}^{m_t} & \xrightarrow{\mathcal{S}_{m_t}^{-1}} & \mathcal{X} \\ & & & & & & & & & & \\ x & \longmapsto & (w_s)_{s=1}^{m_t} & \longmapsto & (y_s)_{s=1}^{t} & \longmapsto & (z_s)_{s=1}^{t} & \longmapsto & (\hat{w}_s(t))_{s=1}^{m_t} & \longmapsto & \hat{x}_t\,. \end{array} \tag{13}$$

In other words, in this scheme we first use the dyadic expansion map to transform x into a string of bits $(w_1, w_2, \ldots, w_{m_t}, \ldots)$ and then we encode the latter into a sequence of channel inputs. The received data are decoded by a block decoder providing an estimated version $(\hat{w}_1(t), \hat{w}_2(t), \ldots, \hat{w}_{m_t}(t))$ of $(w_1, w_2, \ldots, w_{m_t})$ (whose components in general depend on t) which is translated to an estimate $\hat{x}_t$ of x.

2.5 A repetition coding scheme

In this section, tradeoffs between computational complexity and performance of the coding schemes are investigated. First a simple linear-time encodable/decodable scheme is analyzed, showing that the estimation error converges to zero sub-exponentially fast with degree $\alpha = 1/2$. Then, in Sect. 2.6, lower bounds on the estimation error are obtained: It is shown that encoding schemes with finite memory (i.e. that can be implemented by finite-state automata controlled by the dyadic expansion of the source vector), have estimation error bounded away from zero, while finite-window linear-time encodable schemes (i.e. such that each channel input can be written a deterministic function of a finite number of bits of the dyadic expansion of the source vector) cannot achieve a convergence degree larger than $1/2$.

Let $\mathcal{S}_t : \mathcal{X} \to \{0,1\}^t$ be the truncated dyadic expansion map introduced in Sect. 2.4, and let $\mathcal{S}_t^{-1} : \{0,1\}^t \to \mathcal{X}$ be one of its right inverses. If a coding scheme with $\mathcal{E} = \mathcal{S}$ were simply used, i.e. if the bits of the dyadic expansion were directly sent through the channel, then the estimation error Δ_t would not converge to 0 as $t \to \infty$. Indeed, with probability ε the first bit of $\mathcal{S}(x)$ would be lost with no possibility of recovering it. It is therefore necessary to introduce redundancy in order to cope with channel erasures. The simplest way to do that consists in using repetition schemes. Of course, since the different bits of the binary expansion $\mathcal{S}(x)$ require different levels of protection, they need to be repeated with a frequency which is monotonically increasing in their significance.

The encoder we propose here is of the following type: at time t, the bit y_t to be sent through the channel coincides with w_{j_t}, the bit in position j_t of the dyadic expansion $\mathcal{S}(x)$. The encoder in this way will depend on the choice of j_t and fits in the scheme proposed in (12) simply by taking $m_t := \max\{j_1, j_2, \dots, j_t\}$.

In the scheme we propose j_t is selected as follows. Fix a positive real q and define $\tau_0 = 0$ and $\tau_k = \lceil q \rceil + \lceil 2q \rceil + \cdots + \lceil kq \rceil$ for $k \in \mathbb{N}$. Notice that, for any $t \in \mathbb{N}$, there exists a unique k such that $\tau_{k-1} + 1 \leq t \leq \tau_k$. Then, define $j_t := t - \tau_{k-1}$. In other words, we have

$$(y_s)_{s=1}^{\infty} = \tilde{\mathcal{E}}((w_s)_{s=1}^{\infty}) = (w_1, w_2, \dots, w_{\lceil q \rceil}, w_1, w_2, \dots, w_{\lceil 2q \rceil}, w_1, w_2, \dots, w_{\lceil 3q \rceil}, \dots\dots) . \tag{14}$$

In any scheme of this kind the decoding is elementary. The output of the decoder $(\hat{w}_j(t))_{j=1}^{m_t} \in \{0,1\}^{m_t}$ may be given by

$$\hat{w}_j(t) = \begin{cases} z_s & \text{if } \exists s \leq t \text{ such that } j(s) = j \text{ and } z_s \neq ? \\ 0 & \text{otherwise} . \end{cases}$$

Notice that this decoding scheme has complexity growing linearly in t. Indeed, it admits the following natural recursive implementation. First, initialize $\hat{w}_j(0) = 0$ for all $j = 0$. Then, for all $t \geq 0$, upon receiving z_{t+1} we compute $(\hat{w}_j(t+1))_{j=1}^{m_{t+1}}$ as

$$\hat{w}_j(t+1) = \begin{cases} z_{t+1} & \text{if } j = j(t+1) \text{ and } z_{t+1} \neq ? \\ \hat{w}_j(t) & \text{otherwise} \end{cases} . \tag{15}$$

Proposition 1. *Consider the repetition coding scheme defined by (14) and (15) on the BEC with erasure probability ε. Then, the root mean squared error satisfies*

$$\Delta_t \leq p(t) 2^{-\beta t^{1/2}} , \tag{16}$$

where

$$\begin{array}{lll} \beta = \frac{\sqrt{2q}}{d}, & p(t) = C_1 & \text{if } q < \frac{d\log\varepsilon^{-1}}{2} \\ \beta = \frac{\sqrt{2q}}{d}, & p(t) = C_2\sqrt{t} & \text{if } q = \frac{d\log\varepsilon^{-1}}{2} \\ \beta = \frac{\log\varepsilon^{-1}}{\sqrt{2q}}, & p(t) = C_3 & \text{if } q > \frac{d\log\varepsilon^{-1}}{2} \end{array}$$

with C_1, C_2, C_3 positive constants depending only on q, ε and d.

Remark 2. *Notice that Proposition 1 implies that a convergence degree $\alpha = 1/2$ is achievable for any choice of the positive parameter q, without any knowledge of the value of the erasure probability $\varepsilon \in [0, 1[$. If one knows ε, then it is possible to optimize the convergence rate β by choosing $q = \frac{d\log\varepsilon^{-1}}{2}$.*

2.6 A trade-off result between performance and complexity

We shall now show how complexity limitations imply lower bounds to the error decay stronger than Theorem 1. In particular we shall prove that, for certain class of encoders (finite-window and finite-state automata), exponential decay of error can never be achieved. As before, let us assume that $\mathcal{S} : \mathcal{X} \to \{0,1\}^{\mathbb{N}}$ is the dyadic expansion map introduced in Sect. 2.4, and consider encoders $\tilde{\mathcal{E}} : \{0,1\}^{\mathbb{N}} \to \{0,1\}^{\mathbb{N}}$ of the form $\tilde{\mathcal{E}}((w_s)_{s=1}^{\infty})_t = \tilde{E}_t(w_1, \dots, w_{m_t})$, for some finite integer m_t, and a map $\tilde{E}_t : \mathcal{Y}^{m_t} \to \mathcal{Y}$.

In general, $\tilde{E}_t$ may actually depend on a proper subset of the m_t bits $\{1, 2, \dots, m_t\}$. Consider the minimal $\Theta_t \subseteq \{1, 2, \dots, m_t\}$ which allows one to write

$$\tilde{E}_t((w_s)_{s=1}^{m_t}) = f_t((w_s)_{s\in\Theta_t})$$

for a suitable function $f_t : \Theta_t \to \{0,1\}$. Let $n_t = |\Theta_t|$. The encoder $\tilde{\mathcal{E}}$ is called finite-window if n_t is bounded in t. With each encoder it is possible to associate, for every $j, t \in \mathbb{N}$, the quantity $\omega_j(t) := \sum_{1\le s\le t} \mathbf{1}_{\Theta_s}(j)$, counting the number of channel inputs up to time t, which have been affected by w_j. Define

$$\chi_t := \sum\nolimits_{j\in\mathbb{N}} \omega_j(t) = \sum\nolimits_{s\le t} n_s \,.$$

The quantity χ_t is related to the complexity of the encoder $\tilde{\mathcal{E}}$. If the maps f_t are $\mathbb{Z}_2$-linear and separately computed, then χ_t provides an upper bound to the number of binary operations implemented by the encoder up to time t. However, there could be hidden recursive links among the f_t capable to lower the real computational complexity. In any case, for brevity, we shall refer to χ_t as the complexity function of the encoder. The following is our main result, relating the root mean squared error Δ_t to the complexity function χ_t.

Theorem 2. *For any transmission scheme for the BEC, with erasure probability ε, consisting of an encoder with complexity function χ_t, it holds*

$$\Delta_t \geq C\, 2^{-\sqrt{\frac{1}{d}\chi_t \log \varepsilon^{-1}}}\,, \tag{17}$$

where $C > 0$ is a constant depending only on d, the erasure probability ε and the density function f of the random vector x.

We have the following straightforward consequence for finite-window encoders which show that the degree $\alpha = 1/2$ can not be beaten.

Corollary 1. *For any transmission scheme for the BEC, with erasure probability ε, consisting of a finite-window encoder with $n_t \leq n_{\max}$ for every t, it holds*

$$\Delta_t \geq C\, 2^{-\beta t^{1/2}}\,, \tag{18}$$

where $\beta = \sqrt{\frac{n_{\max} \log \varepsilon^{-1}}{d}}$ and where $C > 0$ is a constant depending only on d, the erasure probability ε and the density function f of the random vector x.

Remark 3. *In the case of the repetition encoders treated above, we have that $n_{\max} = 1$. If we compare (18) with (16), we have thus established that among the repetition schemes ($n_{\max} = 1$), the example treated above is optimal from the point of view of the asymptotic performance (both degree and rate of convergence).*

The bound (17) implies that, in order to obtain exponential convergence of the error, χ_t needs to grow at least quadratically in t or, equivalently, that $\frac{1}{t}\chi_t$, i.e. the average number of bits of the dyadic expansion $\mathcal{S}(x)$ the channel inputs depend on, grows at least linearly in t. Indeed, as we shall see, the random linear codes proposed in Sect. 2.8 have exactly this property. However, observe that this does not imply that linear-time encodable schemes cannot attain exponential error decays in any case, since χ_t is, as already noticed, only an upper bound to the complexity of the encoder, intended as the minimum number of operations required by any implementation of the encoder. A possibility would be to consider maps f_t which, despite being not finite-window, can still be computed with bounded complexity in some recursive way. The most obvious choice would be to consider finite-state automata schemes. Unfortunately, such schemes yield very poor performance, as it will be shown in the next subsection. A less simple choice (and which will not pursued here) would be to consider encoders obtained as serial concatenations of finite-window with finite-state automata schemes.

2.7 Finite-state automata encoders

Encoders which can be implemented as finite state automata yield very poor performance. In fact, the root mean squared error Δ_t in this case does not converge to 0 as $t \to +\infty$. Indeed, assume we are given a finite state alphabet A and two maps $\xi : A \times \{0,1\} \to A$, $\rho : A \times \{0,1\} \to \{0,1\}$. Moreover, fix an initial state $a^* \in A$. To the quadruple (A, ξ, ρ, a^*) we can naturally associate an encoder $\tilde{\mathcal{E}}$ as follows. Given $(w_s)_{s=1}^{\infty} \in \{0,1\}^{\mathbb{N}}$, recursively define $(y_s)_{s=1}^{\infty} = \tilde{\mathcal{E}}((w_s)_{s=1}^{\infty})$ by

$$\begin{cases} a_{t+1} &= \xi(a_t, w_t) \quad a_0 = a^* \\ y_t &= \rho(a_t, w_t) \end{cases}$$

Notice that the state updating map ξ together with the initial condition $a_0 = a^*$ yield a sequence of maps $\xi^{(t)} : \{0,1\}^t \to A$ such that $a_{t+1} = \xi^{(t)}(w_1, \dots w_t)$. If we choose $t = t_0$ in such a way that $2^{t_0} > |A|$, the map $\xi^{(t_0)}$ is necessarily not injective. Hence, there exist two different input truncated sequences $(w'_1, \dots, w'_{t_0})$ and $(w''_1, \dots, w''_{t_0})$ such that $\xi^{(t_0)}(w'_1, \dots, w'_{t_0}) = \xi^{(t_0)}(w''_1, \dots, w''_{t_0})$. Consider the event $A = \{w_k = w'_k, \ z_k =? \ \text{for} \ k = 1, \dots, t_0\}$. Clearly, conditioned on A, the decoder, for any $t \geq t_0$, will decode incorrectly at least one information bit in the first t_0 position with positive probability independent from t. Hence, $\Delta_t^2 \geq \mathbb{E}\left[||x - \hat{x}_t||^2 \,|\, A\right] \mathbb{P}(A) \geq 2^{-2t_0/d} \mathbb{P}(A)$.

2.8 A coding scheme with exponential error rates

The goal of this section is to show that, removing the complexity bounds, exponential convergence can be achieved. The proposed scheme will require quadratic computational complexity at the encoder and cubic complexity at the decoder.

We shall use random coding arguments employing linear tree codes over the binary field $\mathbb{Z}_2$. These arguments were first developed in the context of convolutional codes [79, 127], and recently applied in the framework of anytime information theory [116, 117]. For the reader's convenience, and since those results have not appeared anywhere else in this form, we shall present self-contained proofs. The coding strategy we shall propose is very close in spirit to those in [116, Th.5.1] and [117, Th.5.1], the main difference being that we use linear convolutional codes instead of general random convolutional codes. Our choice has the double advantage of lowering the memory and complexity requirements for the encoder and the decoder, and improving the achievable error rate for a significant range of values of ε (see Theorem 4 and Remark 4).

A random causal linear coding scheme In this section we shall identify the binary set $\mathcal{Y} = \{0, 1\}$ with the binary field $\mathbb{Z}_2$ of the integers modulo 2.

Fix a rate $0 < R < 1$ and any t let $m_t := \lfloor Rt \rfloor$. Consider a random, doubly infinite, binary matrix $\phi \in \mathbb{Z}_2^{\mathbb{N}\times\mathbb{N}}$ distributed as follows: $\phi_{ij} = 0$ for all $j > Ri$ (i.e. for all $j \geq m_i + 1$), while $\{\phi_{ij}\}_{1\leq j\leq Ri}$ is a family of mutually independent random variables with identical uniform distribution over $\mathbb{Z}_2$. As customary in random coding arguments, we shall assume the random matrix ϕ to be independent from the source vector x as well as from the channel, and known a priori both at the transmitting and receiving ends. Let us naturally identify the random matrix ϕ with the corresponding random $\mathbb{Z}_2$-linear operator $\tilde{\mathcal{E}} : \mathbb{Z}_2^{\mathbb{N}} \to \mathbb{Z}_2^{\mathbb{N}}$. Consider the truncated encoder

$$\tilde{\mathcal{E}}_t : \mathbb{Z}_2^{m_t} \to \mathbb{Z}_2^t\,, \qquad \tilde{\mathcal{E}}_t\left((w_s)_{s=1}^{m_t}\right) := \pi_t(\phi\mathbf{w})\,, \tag{19}$$

where $\mathbf{w} \in \mathbb{Z}_2^{\mathbb{N}}$ is such that $\pi_{m_t}\mathbf{w} = (w_s)_{s=1}^{m_t}$. Now, let $\mathcal{S} : \mathcal{X} \to \mathbb{Z}_2^{\mathbb{N}}$ be the dyadic expansion map introduced in Sect. 2.4, and define the encoding scheme $\mathcal{E} : \mathcal{X} \to \mathbb{Z}_2^{\mathbb{N}}$ as the composition $\mathcal{E} = \tilde{\mathcal{E}} \circ \mathcal{S}$.

For the decoding part, we shall consider maximum a posteriori decoders $\tilde{\mathcal{D}}_t$. For the special case of the BEC, given the channel outputs z_t, the decoded block at time t, $(\hat{w}_s(t))_{s=1}^{m_t} = \tilde{\mathcal{D}}_t((z_s)_{s=1}^t)$, is defined to be any vector in $\{0,1\}^{m_t}$ which is compatible with the observed channel output $(z_s)_{s=1}^t$. Formally, let $\Xi_t := \{s \in \{1,\dots,t\} : z_s \neq ?\}$ be the set of non-erased positions up to time t, and $\pi_{\Xi_t} : \mathbb{Z}_2^t \to \mathbb{Z}_2^{\Xi_t}$ be the canonical projection. Then, a MAP decoder $\mathcal{D}_t : \{0,1,?\}^t \to \mathbb{Z}_2^{m_t}$ maps the channel output $(z_s)_{s=1}^t$ into any binary string $(\hat{w}_s(t))_{s=1}^{m_t}$ such that

$$\pi_{\Xi_t}\tilde{\mathcal{E}}_t((\hat{w}_s(t))_{s=1}^{m_t}) = \pi_{\Xi_t}(z_s)_{s=1}^t = \pi_{\Xi_t}\tilde{\mathcal{E}}_t((w_s)_{s=1}^{m_t})\,. \tag{20}$$

Finally, the overall decoder is defined as the composition $\mathcal{D}_t := \mathcal{S}_{m_t}^{-1} \circ \tilde{\mathcal{D}}_t$.

Performance analysis We now analyze the coding scheme we have introduced. Notice first of all that, the decoded block $(\hat{w}_s(t))_{s=1}^{m_t} = \tilde{\mathcal{D}}_t((z_s)_{s=1}^t) \in \mathbb{Z}_2^{m_t}$ is uniquely defined, and correct, whenever the linear map $\pi_{\Xi_t}\tilde{\mathcal{E}}_t : \mathbb{Z}_2^{m_t} \to \mathbb{Z}_2^{\Xi_t}$ is injective. However, our analysis requires more detailed information regarding the location of the uncorrectly decoded information bits when injectivity is lost. To this end, let $\{\delta_1, \delta_2, \dots, \delta_{m_t}\}$ be the canonical basis of $\mathbb{Z}_2^{m_t}$, and, for $0 \leq j \leq m_t$, consider the subspace[4] $K_j := \mathrm{span}(\delta_{j+1}, \dots, \delta_{m_t}) \subseteq \mathbb{Z}_2^{m_t}$. For $0 \leq j \leq m_t$,

[4] We shall use the standard convention $\mathrm{span}(\emptyset) := \{0\}$.

define the event $A_j := \{\ker(\pi_{\Xi_t}\tilde{\mathcal{E}}_t) \subseteq K_j\}$. Also, let us define $B_j := A_{j-1} \setminus A_j$, for $1 \leq j \leq m_t$. Observe that $A_j \subseteq A_{j-1}$, and that A_0 coincides with the whole sample space Ω. Hence, for every $t \in \mathbb{N}$, the sample space admits the partition

$$\Omega = \bigcup_{1 \leq j \leq m_t} B_j \bigcup A_{m_t} . \tag{21}$$

Notice now that, from (20) we can deduce that $(w_s - \hat{w}_s(t))_{s=1}^{m_t} \in \ker \pi_{\Xi_t}\tilde{\mathcal{E}}_t$. Therefore, if A_j holds true, then $(\hat{w}_s(t))_{s=1}^{j} = (w_s)_{s=1}^{j}$, i.e. the first j bits of the quantization of x are correctly decoded. We immediately get from (11) that, if A_j occurs, then

$$||\hat{x}_t - x||^2 \leq 4d2^{-2j/d}, \qquad 0 \leq j \leq m_t . \tag{22}$$

The following result characterizes the average mean squared error of the random coding scheme $(\mathcal{E}, \mathcal{D})$ over the BEC. Here the average has to be considered with respect to the randomness of the vector x, the channel, as well as the matrix ϕ. For $\varepsilon \in [0,1]$ and $d \in \mathbb{N}$, define

$$\underline{\beta}'(d, \varepsilon, R) := \min\{\tfrac{1}{d}R, \tfrac{1}{2} \min_{0 \leq \eta \leq 1} D(\eta||1-\varepsilon) + \lfloor \eta - R \rfloor_+\}, \tag{23}$$

where $D(x||y) := x \log \frac{x}{y} + (1-x) \log \frac{1-x}{1-y}$ denotes the binary Kullback-Leiber distance and where $\lfloor x \rfloor_+ := \max\{0, x\}$.

Theorem 3. *Assume transmission over the BEC. Then, for all $0 < R < 1$, the average estimation error of the above-described random coding scheme satisfies*

$$\left(\mathbb{E}||x - \hat{x}_t||^2\right)^{1/2} \leq C\sqrt{t}2^{-\beta'(d,\varepsilon,R)\,t} \tag{24}$$

for all $t \in \mathbb{N}$, where $C > 0$ is a constant depending only on d, R and ε.

Standard probabilistic arguments allow one to prove the following corollary of Theorem 3, characterizing the exponential error rate of a typical realization of the random coding scheme $(\mathcal{E}, \mathcal{D})$. Observe that the root mean squared error of the coding scheme is given by $\left(\mathbb{E}\left[||\hat{x}_t - x||^2|\phi\right]\right)^{1/2}$ which is a function of ϕ, hence a random variable.

Corollary 2. *Assume transmission over the BEC with erasure probability ε. Then, for all $0 < R < 1$, with probability one,*

$$\left(\mathbb{E}[||x - \hat{x}_t||^2|\phi]\right)^{1/2} \leq Ct^{3/2}2^{-\underline{\beta}'(d,\varepsilon,R)\,t}, \tag{25}$$

for a positive constant C.

It is possible to derive another lower bound on the typical-case exponential error rate achieved by the random scheme $(\mathcal{E}, \mathcal{D})$, which turns out to be tighter than that provided by Corollary 2 for certain values of R and ε. For every $0 \le R \le 1$, define $\gamma(R) := \min\{x \in [0,1] : \mathrm{H}(x) \ge 1 - R\}$, and

$$\underline{\beta}''(d,\varepsilon,R) := \min\left\{\frac{1}{d}R,\ \frac{1}{2}\min_{\gamma(R)\le\eta\le 1}\{\mathrm{H}(\eta) - 1 + R - \eta\log\varepsilon\}\right\}.$$

Theorem 4. *Assume transmission over the BEC with erasure probability ε. Then, for all $0 < R < 1$, $\delta > 0$, with probability one*

$$\left(\mathbb{E}[||x - \hat{x}_t||^2|\phi]\right)^{1/2} \le Kt2^{-(\underline{\beta}''(d,\varepsilon,R)+\delta)t}, \tag{26}$$

for a constant $K > 0$.

Remark 4. *It follows from Corollary 2 and Theorem 4 that, for all $R < 1 - \varepsilon$ random causal linear codes achieve exponential convergence rate. Optimizing over $R \in]0, 1 - \varepsilon[$, this shows that the exponent*

$$\underline{\beta}(d,\varepsilon) := \max_{0\le R\le 1}\max\{\underline{\beta}'(d,\varepsilon,R), \underline{\beta}''(d,\varepsilon,R)\}, \tag{27}$$

is achievable. In Fig. 1 the upper and lower bounds to the error exponent, i.e. $\overline{\beta}(d,\varepsilon)$ and $\underline{\beta}(d,\varepsilon)$, are plotted as functions of the erasure probability ε, in the case $d = 1$. Define $\underline{\beta}'(d,\varepsilon) := \max\{\underline{\beta}'(d,\varepsilon,R) : R \in [0,1]\}$, and $\underline{\beta}''(d,\varepsilon) := \max\{\underline{\beta}''(d,\varepsilon,R) : R \in [0,1]\}$. Then, it is not difficult to see that $\lim_{\varepsilon\downarrow 0}\underline{\beta}'(d,\varepsilon) = 1/(d+2)$, while $\lim_{\varepsilon\downarrow 0}\underline{\beta}''(d,\varepsilon) = 1/d$. Hence, Theorem 4 becomes particularly relevant for small erasure probabilities, showing that the noiseless error exponent $1/d$ (see Sect. 2.4) is recovered in the limit of vanishing noise: this does not follow from the average-code analysis of Theorem 1.

Computational complexity of the scheme Observe that the number n_t of binary operations required in order to compute the channel input $y_t = \tilde{\mathcal{E}}_t((w_s)_{s=1}^{m_t})$, equals the number of non-zero entries of the t-th row of the infinite random matrix ϕ. By the way ϕ has been defined, n_t is a binomial random variable of parameters m_t and $1/2$. Hence, the number of binary operations required by the encoder up to time t, $\chi_t := \sum_{s\le t} n_s$, has binomial distribution of parameters $\frac{1}{2}m_t(m_t+1)$ and $1/2$. Therefore, the worst-case encoding complexity (worst case with respect to the realization of ϕ) grows like $\frac{1}{2}R^2t^2$, while the strong law of large numbers

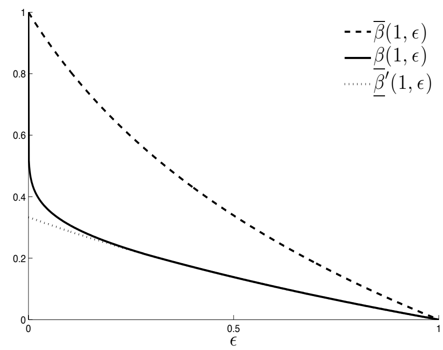


Figure 1: Upper and lower bounds to the achievable estimation error exponent achievable on the BEC (as defined in (9), (27) and Remark 4, respectively) are plotted as a function of the erasure probability ε for $d = 1$.

implies that the typical encoder complexity χ_t is such that $\chi_t/\frac{1}{4}R^2t^2$ converges to 1 with probability one. Thus, the encoder complexity (both worst-case and typical-case) is quadratic in t. Further, observe that the memory requirements of the encoder are quadratic in t for it is necessary to store $m_t t$ binary values in order to memorize the finite truncation $\mathcal{E}_t$ of the encoder $\mathcal{E}$.

In order to evaluate the decoder's computational complexity, observe that $\tilde{\mathcal{D}}_t$ is required to solve the $\mathbb{Z}_2$-linear system

$$\pi_{\Xi_t}\tilde{\mathcal{E}}_t((w_s)_{s=1}^{m_t}) = \pi_{\Xi_t}(z_s)_{s=1}^{t}\,. \tag{28}$$

at each time step t. This can be performed using Gaussian elimination techniques in order to reduce the matrix $\pi_{\Xi_t}\tilde{\mathcal{E}}_t$ to a lower-diagonal form. Notice that a sequential implementation is possible, i.e. the part of $\pi_{\Xi_t}\tilde{\mathcal{E}}_t$ which has been reduced in lower triangular form at time t does not require to be further processed in future times $s > t$. Since Gaussian elimination techniques require $O(t^3)$ operations, we can conclude that the decoder complexity is at most $O(t^3)$. On the other hand, it might be possible to find algorithms for solving a linear system like (28) with number of operations $o(t^3)$: see [125, pagg.247-248] for the analogous problem for linear systems over the reals. However, the system (28) cannot be solved using fewer operations than those required to verify that a given string $v \in \mathbb{Z}_2^{m_t}$ is a solution. Using arguments similar to those outlined above, it is possible to show that, with probability one, this requires $\Theta(t^2)$ binary operations. In summary, the complexity of maximum a posteriori decoding of linear convolutional codes on the BEC is at most $O(t^3)$ and at least $\Theta(t^2)$.

2.9 Simulation results for finite-window coding schemes

We shall now present Monte Carlo simulation results for some finite-window $\mathbb{Z}_2$-linear coding schemes with low-complexity iterative decoding. These schemes are based on ideas similar to those of digital fountain codes (see [99] [104, Ch.50]). The latter are widely used in many applications, such as data storage, or reliable transmission on broadcast channels with erasures. The main additional challenge posed by our application consists in providing unequal error protection to the source bits.

We propose the following random construction for finite-window encoders fitting in the framework of Sect. 2.5. As usual, assume that we have a dyadic expansion $\mathcal{S}$ mapping the vector x into an infinite string of bits $(w_s)_{s=1}^{\infty}$. We imagine that at each time t the encoder produces a bit y_t which is the (modulo-2) sum of a random number of randomly chosen w_s, namely

$$y_t = \sum_{s\in\Theta_t} w_s \,.$$

where Θ_t is a random subset of $\mathbb{N}$. We assume that the cardinality of Θ_t is bounded, i.e. $|\Theta_t| \leq n_{\max}$.

More precisely, fix $n_{\max} \in \mathbb{N}$, and a probability distribution $\mu(\,\cdot\,)$ on $\{1, \dots, n_{\max}\}$. Randomly generate a sequence $(n_t)_{t\in\mathbb{N}}$ of independent random variables distributed according to $\mu(\,\cdot\,)$. Let $(\nu_t(\,\cdot\,))_{t\in\mathbb{N}}$ be a sequence of probability distributions over $\mathbb{N}$, with $\nu_t(\,\cdot\,)$ possibly depending on $(n_s)_{s\leq t}$. Then, for every $t \geq 1$, consider the random set $\Theta_t := \{\theta_{1,t}, \theta_{2,t}, \dots, \theta_{n_t,t}\}$, where $\theta_{i,t}$ are independent random variables uniformly distributed according to $\nu_t(\,\cdot\,)$. Notice that in this way we have that $|\Theta_t| \leq n_t \leq n_{\max}$ and so the encoder complexity is linear in t.

For the decoding, a sequential implementation of the peeling algorithm is used, this being the standard decoding technique for digital fountain codes [99] [104, Ch.50]. Such an algorithm works on an iteratively updated infinite hypergraph[5] $\mathcal{G}_t = (\mathcal{V}_t, \mathcal{H}_t)$ as explained below. At $t = 0$, $\mathcal{G}_0$ is initialized with vertex set $\mathcal{V}_0 = \mathbb{N}$ and empty hyperedge set $\mathcal{H}_0 = \emptyset$. The estimates $(\hat{w}_s(0))_{s\in\mathbb{N}}$ of the dyadic expansion $\mathcal{S}(x)$ are in turn initialized arbitrarily in $\{0,1\}^{\mathbb{N}}$. At each time $t \geq 1$, first update $\mathcal{V}_t = \mathcal{V}_{t-1}$, $\mathcal{H}_t = \mathcal{H}_{t-1}$, and $\hat{w}_s(t) = \hat{w}_s(t+1)$ for all $s \in \mathbb{N}$. Then:

[5] The term hypergraph [57, pag.7] refers to a pair $(\mathcal{V}, \mathcal{H})$, where $\mathcal{V}$ is a discrete set and $\mathcal{H}$ is a subset of $\mathcal{P}(\mathcal{V})$, the power set of $\mathcal{V}$.

- if $z_t = ?$, then quit; if $z_t \neq ?$, update $\mathcal{H}_t = \mathcal{H}_t \cup \{B_t\}$, where $B_t := \Theta_t \cap \mathcal{V}_t$;
- if $|B_t| > 1$, then quit; otherwise if $B_t = \{v\}$ for some $v \in \mathcal{V}_t$, set $\hat{w}_v(t) = z_t + \sum_{j \in \Theta_t \setminus \{v\}} \hat{w}_j(t)$, eliminate v from $\mathcal{V}_t$ as well as from all the hyperedges $h \in \mathcal{H}_t$ containing it;
- if $|h| \neq 1$ for all $h \in \mathcal{H}_t$, quit; otherwise, if there is some $h = \{v\} \in \mathcal{H}_t$, repeat the previous step.

The above-described algorithm requires an order of $\chi_t = \sum_{s \leq t} n_s$ operations up to time t, hence it has linear complexity in t. It is suboptimal with respect to the maximum a posteriori decoding: it may fail to correctly estimate the first j bits of the dyadic expansion $\mathcal{S}(x)$ even when that would be possible using the maximum a posteriori decoder.

In Fig. 2 we report Monte Carlo simulations of three finite-windows encoding schemes, with $n_{\max} = 1, 2, 4$ respectively. The degree distribution $\mu(\,\cdot\,)$ was chosen to be the truncated soliton one [104, pag.592]

$$\mu(1) := \frac{1}{n_{\max}}\,, \qquad \mu(n) := \frac{1}{n(n-1)} \quad \forall\, 2 \leq n \leq n_{\max}\,. \tag{29}$$

The distributions ν_t have been selected as follows. We define $\rho := 2\left(d \log \varepsilon^{-1}\right)^{-1}$, $s_t := \lfloor \sqrt{2\chi_t \rho^{-1}} \rfloor$, and $\varsigma_t = \frac{\chi_t}{s_t} + \rho \frac{s_t + 1}{2}$, where $\chi_t = \sum_{s \leq t} n_s$. Then choose

$$\nu_t(j) := \begin{cases} \eta(\varsigma_t - \rho j) & \text{if} \quad j \leq s_t \\ 0 & \text{if} \quad j > s_t\,, \end{cases} \tag{30}$$

It is clear from Fig. 2(a) that the three schemes have subexponential error decay and that increasing the degree allows one to obtain better convergence rates. Fig. 2(b) shows that the convergence degree is $\alpha = 1/2$, as expected from the theory, while it is possible to recognize the different values of β of the three schemes, in the asymptotic limit of $-\frac{1}{\sqrt{t}} \log \Delta_t$.

It should be underlined as the choices of the distributions μ and ν_t were not optimized, but rather suggested by the literature on digital fountain codes and by Theorem 2, respectively. A theoretical analysis of the behavior of finite-window schemes, hopefully providing hints on the design of μ and ν_t, is left as a topic for future research.

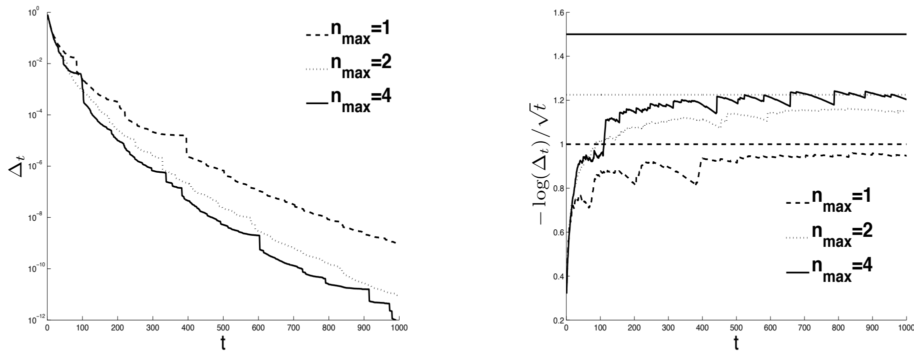


Figure 2: Monte Carlo simulations of finite-window coding schemes on the BEC, with erasure probability $\varepsilon = 0.5$. The performance of three coding schemes are compared: these schemes were randomly generated accordingly to (29) and (30) with $n_{\max} = 1, 2, 4$ respectively. In (a) the root mean squared error Δ_t is plotted as a function of the time t in log-linear scale. In (b) $-\frac{1}{\sqrt{t}} \log \Delta_t$ is plotted as a function of t, together with the corresponding upper bounds $\sqrt{\chi_t \log \varepsilon^{-1}}$ provided by Theorem 2. The number of samples used is 200000.

3 Stabilization Over Gaussian Sensor Networks

In this section we study the problem of remotely stabilizing a discrete first order LTI system over a Gaussian sensor network. The sensor network channel consists of one sender (source), one receiver (destination) and a number of intermediate nodes (relays) whose sole purpose is to help the communication between the source and the destination. The achievable information rate over a sensor network channel depends on the processing strategy of the sensor (relay) nodes. The most well known relaying strategies are amplify-and-forward (AF), compress-and-forward, and decode-and-forward [94]. AF strategy is well suited for delay sensitive closed-loop control applications and is therefore addressed in this work. For communication and control we propose to use the Schalkwijk-Kailath based coding strategy [62] which is suitable for channels with feedback [85, 118]. We used the Schalkwijk-Kailath based scheme to obtain stability regions for control over multiple-access, broadcast, and interference channels in [8, 12]. In [1, 9] we derived rate sufficient conditions for stabilization of an LTI plant over Gaussian relay channels. In this report we present extension of our previous works to Gaussian sensor networks where a large number of relay nodes are deployed to communicate the state process to the remote controller. The objective of this work is to derive sufficient conditions for stability of an LTI plant in mean square sense [24, 82, 108, 109, 116, 120] over some fundamental topologies of sensor networks.

Some of the results presented in this section have appeared in [1, 9] and a journal article [41] based on this work is in preparation.

3.1 Relevant literature

The problem of remotely controlling dynamical systems over communication channels has gained significant attention in recent years. Such problems ask for interaction between stochastic control theory and information theory [44,53]. The minimum data rate below which the stability of an LTI system is impossible has been derived in stochastic and deterministic settings in [44, 108, 109, 124], where they considered quantization errors and noise-free rate-limited channels. In [102, 123] are necessary rate conditions required to stabilize an LTI plant almost surely. However, from [116] we know that the characterization by Shannon capacity is not enough for sufficient conditions for moment stability in closed-loop control. In [120] a simple coding scheme is proposed to mean square stabilize an LTI plant over noise-free rate-limited channels. The mean square stability of discrete plant over signal-to-noise ratio constraint channels is addressed in [29,40,42,105]. In [16] the authors considered noisy communication links between both observer–controller and controller–plant. In [24] the necessary and sufficient conditions are derived for mean square stability of an LTI system over time varying feedback channels.

3.2 Problem Formulation

We consider a scalar discrete-time LTI system, whose state equation is given by

$$X_{t+1} \quad = \quad \lambda X_t + U_t + W_t, \tag{31}$$

where $\{X_t\} \subseteq \mathbb{R}$, $\{U_t\} \subseteq \mathbb{R}$, and $\{W_t\} \subseteq \mathbb{R}$ are state, control, and plant noise processes. The plant noise $\{W_t\}$ is a zero mean white Gaussian noise sequence with variance n_w. We assume that the open-loop system is unstable ($|\lambda| > 1$) and the initial state X_0 is a random variable with variance α_0 and an arbitrary probability distribution. We consider a remote control setup, where the observed state value is transmitted to a controller over a Gaussian sensor network as shown

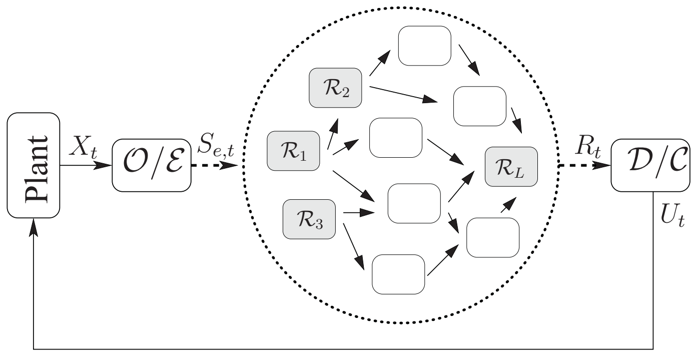


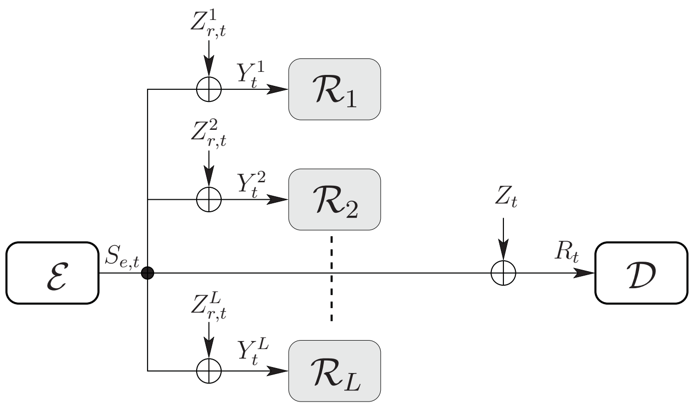
Figure 3: The unstable plant has to be controlled by the actions of observer/encoder ($\mathcal{O}/\mathcal{E}$) and decoder/controller ($\mathcal{D}/\mathcal{C}$) over a sensor (relay) network.

in Fig. 3. In order to communicate the observed state value X_t over the noisy sensor (relay) network, an encoder $\mathcal{E}$ is lumped with the observer $\mathcal{O}$ and a decoder $\mathcal{D}$ is lumped with the controller $\mathcal{C}$. In addition there are L sensor (relay) nodes $\{\mathcal{R}_i\}_{i=1}^{L}$ within the channel to support communication from $\mathcal{E}$ to $\mathcal{D}$. At any time instant t, $S_{e,t}$ and R_t are the input and the output of the sensor (relay) network and U_t is the control action. Let f_t denote the observer/encoder policy such that $S_{e,t} = f_t(X_0^t, U_0^{t-1})$, where $X_0^t = \{X_s, 0 \leq s \leq t\}$ and we have the following average power constraint $\lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}[S_{e,t}^2] \leq P_S$. Further let γ_t denote the decoder/controller policy, then $U_t = \gamma_t(R_0^t)$. The objective in this paper is to find conditions on the system parameters so that the plant in (70) can be mean square stabilized over a given Gaussian sensor network.

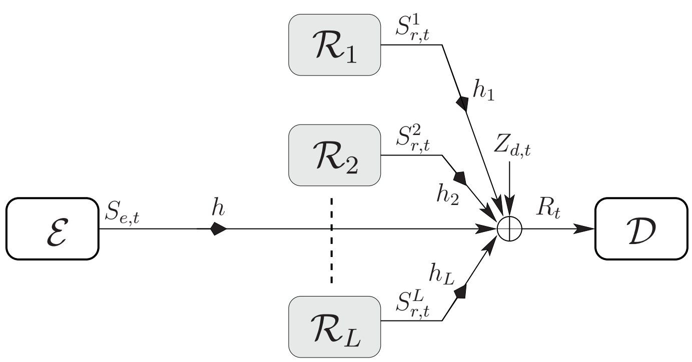
Definition 1. *A system is said to be* mean square stable *if there exists a constant $M < \infty$ so that $\mathbb{E}[X_t^2] < M$ for all t.*

A general sensor network consists of an arbitrary number of sensor nodes with arbitrary communication links. In order to understand the problem of stabilization over a general network, we study some basic network topologies such as non-orthogonal sensor network, cascade sensor network, and parallel sensor network. These topologies serve as the basic building blocks of a large sensor network. In practice a sensor nodes can be either full-duplex[6] or half-duplex, we therefore

[6] A full-duplex sensor node is capable of transmitting and receiving signals simultaneously



(a) First transmission phase.



(b) Second transmission phase.

study both configurations. We also briefly discuss the scenarios when the state encoder or the remote controller are equipped with multiple antennas to provide spatial diversity for communication. For mean square stabilizing the first order linear system in (70) over these basic sensor network settings we present some sufficient and necessary conditions and some interesting insights on the optimal schemes.

3.3 Non-orthogonal Half-duplex Sensor Network

A general non-orthogonal half-duplex Gaussian sensor (relay) network with L sensor (relay) nodes $\{\mathcal{R}_i\}_{i=1}^{L}$ is illustrated in Fig. 3.3. By non-orthogonal network we mean that all the communicating nodes share a common signal space, that is they transmit signals in overlapping time slots and same frequency bands.The variables $S_{e,t}$ and $S_{r,t}^i$ denote the transmitted signals from the state encoder $\mathcal{E}$ and the sensor $\mathcal{R}_i$ at any discrete time step t. The variables $Z_{r,t}^i$ and $Z_{d,t}$ denote the mutually independent white Gaussian noise components at the sensor node i and at the decoder $\mathcal{D}$ of the remote control unit respectively, with $Z_{r,t}^i \sim \mathcal{N}(0, N_r^i)$ and $Z_{d,t} \sim \mathcal{N}(0, N_d)$. The noise components $\{Z_{r,t}^i\}_{i=1}^{L}$ are independent across the sensors, i.e., $\mathbb{E}[Z_{r,t}^k Z_{r,t}^i] = 0$ for all $i \neq k$. The information transmission from the state encoder consists of two phases as shown in Fig. 3.3. In the first transmission phase the encoder $\mathcal{E}$ transmits a signal with an average power $2\beta P_S$, where $0 < \beta \leq 1$ is a parameter that allocates power to the two transmission phases. In this transmission phase all the sensor nodes $\{\mathcal{R}\}_{i=1}^{L}$ listen but remain silent. In the second transmission phase, the encoder $\mathcal{E}$ and the sensor nodes $\{\mathcal{R}_i\}_{i=1}^{L}$ transmit simultaneously. In this second transmission phase, the encoder transmits with an average power $2(1-\beta)P_S$ and the i-th sensor node transmits with an average power $\mathbb{E}[(S_{r,t}^i)^2] = P_R^i$ such that $\sum_{i=1}^{L} P_R^i \leq P_R$. In this section we consider a Accordingly the output of the sensor network at the decoder (controller) is R_t which is given by

$$\begin{aligned} R_t &= hS_{e,t} + Z_t && t = 1, 3, 5, \ldots \\ R_t &= hS_{e,t} + \sum_{i=1}^{L} h_i S_{r,t}^i + Z_t, && t = 2, 4, 6, \ldots \end{aligned}$$

where $h \in \mathbb{R}$ denotes the gain of $\mathcal{E} - \mathcal{D}$ link and $h_i \in \mathbb{R}$ denotes the gain of $\mathcal{R}_i - \mathcal{D}$ link.

We now present a sufficient condition for the mean square stability of the first order linear system in (70) over the given non-orthogonal half-duplex sensor network.

Theorem 5. *The scalar linear time invariant system in (70) can be mean square*

while a half-duplex sensor node cannot simultaneously receive and transmit signals.

stabilized over the non-orthogonal half-duplex *sensor network if*

$$\log(\lambda) < \frac{1}{4} \max_{\substack{0<\beta\leq 1 \\ P_R^i : \sum_i P_R^i \leq P_R}} \left\{ \log\left(1 + \frac{2h^2\beta P_S}{N_d}\right) + \log\left(1 + \frac{\tilde{M}\left(\beta, \{P_R^i\}_{i=1}^L\right)}{\tilde{N}\left(\beta, \{P_R^i\}_{i=1}^L\right)}\right)\right\}, \tag{32}$$

where $\beta \in [0,1]$, $\tilde{M}\left(\beta, \{P_R^i\}_{i=1}^L\right) = \left(\sqrt{2h^2(1-\beta)P_S} + \sqrt{\frac{2\beta P_S N_d}{(2h^2\beta P_S + N_d)}}\left(\sum_{i=1}^L \sqrt{\frac{h_i^2 P_R^i}{2\beta P_S + N_R^i}}\right)\right)^2$, *and* $\tilde{N}(\beta, \{P_R^i\}_{i=1}^L) = \sum_{i=1}^L \frac{h_i^2 P_R^i N_R^i}{2\beta P_S + N_R^i} + N_d$.

Proof. In order to prove Theorem 5 we propose a linear and memoryless communication and control scheme. This scheme is based on the well-known Schalkwijkn-Kailath coding scheme [62, 118]. By employing the proposed linear scheme over the given *non-orthogonal half-duplex* sensor network, we then find conditions on the system parameters λ which are sufficient to mean square stabilize the system in (70). The control and communication scheme works as follows.

Initial time step, $t = 0$

At time step $t = 0$, the state encoder $\mathcal{E}$ observes X_0 and it transmits $S_{e,0} = \sqrt{\frac{P_S}{\alpha_0}} X_0$. The decoder $\mathcal{D}$ receives $R_0 = hS_{e,0} + Z_{d,0}$. It then estimates X_0 as

$$\hat{X}_0 = \frac{1}{h}\sqrt{\frac{\alpha_0}{P_S}} R_0 = X_0 + \frac{1}{h}\sqrt{\frac{\alpha_0}{P_S}} Z_{d,0}.$$

The controller $\mathcal{C}$ then takes an action $U_0 = -\lambda \hat{X}_0$ which results in

$$\begin{aligned} X_1 &= \lambda X_0 + U_0 + W_0 \\ &= \lambda\left(X_0 - \hat{X}_0\right) + W_0 = -\frac{\lambda}{h}\sqrt{\frac{\alpha_0}{P_S}} Z_{d,0} + W_0. \end{aligned} \tag{33}$$

The new plant state $X_1 \sim \mathcal{N}(0, \alpha_1)$ where $\alpha_1 = \frac{\lambda^2 N_d}{h^2 P_S}\alpha_0 + n_w$.

First transmission phase, $t = 1, 3, 5, \ldots$

The state encoder $\mathcal{E}$ observes X_t and it then transmits $S_{e,t} = \sqrt{\frac{2\beta P_S}{\alpha_t}} X_t$ to the sensor network. The sensor nodes in the network $\{\mathcal{R}_i\}_{i=1}^L$ choose to receive this signal over the Gaussian links and do not transmit any signal in this transmission phase because they are half-duplex. The decoder $\mathcal{D}$ observes $R_t = hS_{e,t} + Z_{d,t}$ and computes the MMSE estimate of X_t, which is given by

$$\begin{aligned}\hat{X}_t &= \mathbb{E}[X_t|R_1, R_2, ..., R_t] \\ &\overset{(a)}{=} \mathbb{E}[X_t|R_t] \overset{(b)}{=} \frac{\mathbb{E}[X_t R_t]}{\mathbb{E}[R_t^2]} R_t \overset{(c)}{=} \left(\frac{h\sqrt{2\beta P_S \alpha_t}}{2h^2\beta P_S + N_d}\right) R_t,\end{aligned}$$

where (a) follows from the *orthogonality principle* of MMSE estimation (that is $\mathbb{E}[X_t R_{t-j}] = 0$ for $j \geq 1$) [34]; (b) follows from the fact that the optimum MMSE estimator for a Gaussian variable is linear [34]; and (c) follows from $\mathbb{E}[X_t R_t] = \sqrt{2h^2\beta P_S \alpha_t}$ and $\mathbb{E}[R_t^2] = 2h^2\beta P_S + N_d$.
The controller $\mathcal{C}$ takes an action $U_t = -\lambda \hat{X}_t$ which results in $X_{t+1} = \lambda(X_t - \hat{X}_t) + W_t$. The new plant state X_{t+1} is linear combination of zero mean Gaussian variables $\{X_t, \hat{X}_t, W_t\}$, therefore it is also zero mean Gaussian with the following variance

$$\begin{aligned}\alpha_{t+1} \triangleq \mathbb{E}[X_{t+1}^2] &= \lambda^2 \mathbb{E}[(X_t - \hat{X}_t)^2] + \mathbb{E}[W_t^2] \\ &= \lambda^2 \left(\frac{N_d}{2h^2\beta P_S + N_d}\right) \alpha_t + n_w, \end{aligned} \tag{34}$$

where the last equality follows from $\mathbb{E}[X_t \hat{X}_t] = \mathbb{E}[\hat{X}_t^2] = \frac{2h^2\beta P_S \alpha_t}{2h^2\beta P_S + N_d}$ (by computation).

Second transmission phase, $t = 2, 4, 6, ...$

The encoder $\mathcal{E}$ observes X_t and it then inputs $S_{e,t} = \sqrt{\frac{2(1-\beta)P_S}{\alpha_t}} X_t$ to the sensor network channel. In this phase the sensor nodes choose to transmit their own signal to the decoder $\mathcal{D}$ and thus they can not listen to the signal transmitted from the state encoder due to their half-duplex nature. Each sensor nodes amplifies the signal that it had received in the previous time step (first transmission phase) under an average power constraint and transmits it to the decoder $\mathcal{D}$. That is all the sensor nodes work as amplify-and-forward relays. The signal transmitted from the i-th sensor node is thus given by, $S_{r,t}^i = \sqrt{\frac{P_R^i}{\left(2\beta P_S + N_R^i\right)}} \left(S_{e,t-1} + Z_{r,t-1}^i\right)$. The decoder $\mathcal{D}$ accordingly receives

$$R_t = hS_{e,t} + \sum_{i=1}^{L} h_i S_{r,t}^i + Z_t = L_1 X_t + L_2 X_{t-1} + \tilde{Z}_t, \tag{35}$$

where $L_1 = \sqrt{\frac{2(1-\beta)h^2 P_S}{\alpha_t}}$, $L_2 = \sum_{i=1}^{L} \sqrt{\frac{2\beta h_i^2 P_S P_R^i}{(2\beta P_S + N_R^i)\alpha_{t-1}}}$, and $\tilde{Z}_t = Z_{d,t} + \sum_{i=1}^{L} \sqrt{\frac{h_i^2 P_R^i}{2\beta P_S + N_R^i}} Z_{r,t-1}^i$ is a white Gaussian noise sequence with zero mean and variance $\tilde{N}(\beta, \{P_R^i\}_{i=1}^L) = N_d + \sum_{i=1}^{L} \frac{h_i^2 P_R^i N_R^i}{2\beta P_S + N_R^i}$. The decoder then computes the MMSE estimate of X_t given all previous channel outputs $\{R_0, R_1, ..., R_t\}$ in the following three steps:

1. Compute the MMSE prediction of R_t from $\{R_1, R_2, ..., R_{t-1}\}$, which is given by

$$\hat{R}_t = L_2 \hat{X}_{t-1},$$

   where $\hat{X}_{t-1}$ is the MMSE estimate of X_{t-1}.

2. Compute the innovation

$$\begin{aligned} I_t = R_t - \hat{R}_t &= L_1 X_t + L_2(X_{t-1} - \hat{X}_{t-1}) + \tilde{Z}_t \\ &\overset{(a)}{=} L_1 X_t + \frac{L_2}{\lambda}(X_t - W_{t-1}) + \tilde{Z}_t = \\ &\left(\frac{\lambda L_1 + L_2}{\lambda}\right) X_t - \frac{L_2}{\lambda} W_{t-1} + \tilde{Z}_t, \end{aligned} \tag{36}$$

   where (a) follows from $X_t = \lambda\left(X_{t-1} - \hat{X}_{t-1}\right) + W_{t-1}$.

3. Compute the MMSE estimate of X_t given $\{R_1, R_2, ..., R_{t-1}, I_t\}$. The state X_t is independent of $\{R_1, R_2, ..., R_{t-1}\}$, therefore we can compute the estimate $\hat{X}_t$ based on I_t only without any loss of optimality, that is

$$\begin{aligned} \hat{X}_t = \mathbb{E}[X_t | I_t] &\overset{(a)}{=} \frac{\mathbb{E}[X_t I_t]}{\mathbb{E}[I_t^2]} I_t \\ &\overset{(b)}{=} \frac{\lambda\left(\lambda L_1 + L_2\right)\alpha_t}{\left(\lambda L_1 + L_2\right)^2 \alpha_t + L_2^2 n_w + \lambda^2 \tilde{N}(\beta, P_R)} I_t, \end{aligned} \tag{37}$$

   where (a) follows from an MMSE estimation of a Gaussian variable; and (b) follows from $\mathbb{E}[X_t I_t] = \left(\frac{\lambda L_1 + L_2}{\lambda}\right)\alpha_t$ and $\mathbb{E}[I_t^2] = \left(\frac{\lambda L_1 + L_2}{\lambda}\right)^2 \alpha_t + \frac{L_2^2 n_w}{\lambda^2} + \tilde{N}(\beta, P_R)$.

The controller $\mathcal{C}$ takes an action $U_t = -\lambda \hat{X}_t$ which results in $X_{t+1} = \lambda(X_t - \hat{X}_t) + W_t$. The new plant state X_{t+1} is linear combination of zero mean Gaussian variables $\{X_t, \hat{X}_t, W_t\}$, therefore it is also zero mean Gaussian with the following variance

$$\begin{aligned}
\alpha_{t+1} &\triangleq \mathbb{E}[X_{t+1}^2] = \lambda^2 \mathbb{E}[(X_t - \hat{X}_t)^2] + \mathbb{E}[W_t^2] \\
&\overset{(a)}{=} \lambda^2 \alpha_t \left(\frac{L_2^2 n_w + \lambda^2 \tilde{N}(\beta, P_R)}{(\lambda L_1 + L_2)^2 \alpha_t + L_2^2 n_w + \lambda^2 \tilde{N}(\beta, P_R)} \right) + n_w \qquad (38) \\
&\overset{(b)}{=} \lambda^2 \left(\lambda^2 k \alpha_{t-1} + n_w \right) \times \\
&\left(\frac{(n_w k_1) \frac{1}{\alpha_{t-1}} + \lambda^2 \tilde{N}(\beta, P_R)}{\left(\lambda k_2 + \sqrt{\frac{k_1}{\lambda^2} (\lambda^2 k + n_w \frac{1}{\alpha_{t-1}})} \right)^2 + (n_w k_1) \frac{1}{\alpha_{t-1}} + \lambda^2 \tilde{N}(\beta, P_R)} \right) + n_w, \\
&= \lambda^2 \left(\lambda^2 k \alpha_{t-1} + n_w \right) \times \\
&\left(\frac{\left(\frac{n_w k_1}{\lambda^2} \right) \frac{1}{\alpha_{t-1}} + \tilde{N}(\beta, P_R)}{\left(k_2 + \sqrt{k_1 k + \frac{n_w k_1}{\lambda^2} \frac{1}{\alpha_{t-1}}} \right)^2 + \left(\frac{n_w k_1}{\lambda^2} \right) \frac{1}{\alpha_{t-1}} + \tilde{N}(\beta, P_R)} \right) + n_w, \qquad (39)
\end{aligned}$$

where (a) follows from $\mathbb{E}[X_t \hat{X}_t] = \mathbb{E}[\hat{X}_t^2] = \frac{(\lambda L_1 + L_2)^2 \alpha_t}{(\lambda L_1 + L_2)^2 \alpha_t + L_2^2 n_w + \lambda^2 \tilde{N}(\beta, P_R)}$; (b) follows by substituting the values of L_1 and L_2; and by substituting $\frac{\alpha_t}{\alpha_{t-1}}$ from (34) and by defining $k \triangleq \frac{N}{2h^2 \beta P_S + N}$, $k_1 \triangleq \frac{2\beta P_S P_R}{2\beta P_S + N_R}$, $k_2 \triangleq \sqrt{2h^2(1 - \beta P_S)}$.

We want to find the values of the system parameter λ for which the second moment of the state remains bounded, i.e., the sequence $\{\alpha_t\}$ has to be bounded. Rewriting (34) and (39), the variance of the state at any time t is given by

$$\alpha_t = \lambda^2 \left(\frac{N}{2h^2 \beta P_S + N} \right) \alpha_{t-1} + n_w, \qquad t = 2, 4, 6, ... \qquad (40)$$

$$\begin{aligned}
\alpha_t &= \lambda^2 \left(\lambda^2 k \alpha_{t-2} + n_w \right) \times \\
&\underbrace{\left(\frac{\left(\frac{n_w k_1}{\lambda^2} \right) \frac{1}{\alpha_{t-2}} + \tilde{N}(\beta, P_R)}{\left(k_2 + \sqrt{k_1 k + \frac{n_w k_1}{\lambda^2} \frac{1}{\alpha_{t-2}}} \right)^2 + \left(\frac{n_w k_1}{\lambda^2} \right) \frac{1}{\alpha_{t-2}} + \tilde{N}(\beta, P_R)} \right)}_{\triangleq f(\alpha_{t-2})} + n_w \\
&= \lambda^2 \left(\lambda^2 k \alpha_{t-2} + n_w \right) f(\alpha_{t-2}) + n_w, \qquad t = 3, 5, 7, ... \qquad (41)
\end{aligned}$$

where $\alpha_1 = \frac{\lambda^2 N}{h^2 P_S} \alpha_0 + n_w$. If the odd indexed sub-sequence $\{\alpha_{2t+1}\}$ in (41) is bounded, then the even indexed sub-sequence $\{\alpha_{2t}\}$ in (40) is also bounded. Therefore it is sufficient to consider the odd indexed sub-sequence $\{\alpha_{2t+1}\}$. We will now construct a sequence $\{\alpha'_t\}$ which upper bounds the sub-sequence $\{\alpha_{2t+1}\}$. Then we will derive conditions on the system parameter λ for which the sequence $\{\alpha'_t\}$ stays bounded and consequently the boundedness of $\{\alpha_{2t+1}\}$ will be guaranteed. In order to construct the upper sequence $\{\alpha'_t\}$, we work on the term $f(\alpha_{t-2})$ in (41) and make use of the following lemma.

Lemma 1. *Consider a function* $f(x) = \frac{a+\frac{b}{x}}{(c+\sqrt{d+\frac{b}{x}})^2+a+\frac{b}{x}}$ *defined in the interval* $[0,\infty)$, *where* $a,b,c,d \geq 0$. *The function* $f(x)$ *can be upper bounded as* $f(x) \leq f_\infty + \frac{m}{x}$ *for some* $m > 0$, *where* $f_\infty \triangleq \lim_{x\to\infty} f(x) = \frac{a}{(c+\sqrt{d})^2+a}$.

Proof. The proof can be found in [9, Appendix C]. ∎

Starting from (41) and by using the above lemma, we write the following series of inequalities

$$\begin{aligned}
\alpha_t &= \lambda^2\left(\lambda^2 k\alpha_{t-2} + n_w\right) f(\alpha_{t-2}) + n_w \\
&\overset{(a)}{\leq} \lambda^2\left(\lambda^2 k\alpha_{t-2} + n_w\right)\left(f_\infty + \frac{m}{\alpha_{t-2}}\right) + n_w \\
&= \lambda^4 k f_\infty \alpha_{t-2} + \frac{\lambda^2 n_w m}{\alpha_{t-2}} + n_w f_\infty + \lambda^4 mk + n_w \\
&\overset{(b)}{\leq} \lambda^4 k f_\infty \alpha_{t-2} + \lambda^2 m + n_w f_\infty + \lambda^4 mk + n_w \triangleq g(\alpha_{t-2}),
\end{aligned} \tag{42}$$

where (a) follows from Lemma 1 and $f_\infty \triangleq \lim_{\alpha\to\infty} f(\alpha) = \left(\frac{\tilde{N}(\beta,P_R)}{(k_2+\sqrt{k_1 k})^2+\tilde{N}(\beta,P_R)}\right)$; and (b) follows from the fact that $\alpha_t \geq n_w$ for all t, which is obvious from (40) and (41) that the value of α_t can never be less than n_w. Since $g(\alpha)$ in (42) is a linearly increasing function, it can be used to construct the sequence $\{\alpha'_t\}$ which upper bounds the odd indexed sub-sequence $\{\alpha_{2t+1}\}$ given in (41). We construct the sequence $\{\alpha'_t\}$ as

$$\begin{aligned}
\alpha_{2t+1} &\leq \alpha'_{t+1} = g(\alpha'_t), \qquad \text{for all } t \geq 1 \\
&\overset{(a)}{=} \lambda^4 k f_\infty \alpha'_t + \lambda^2 m + n_w f_\infty + \lambda^4 mk + n_w \\
&\overset{(b)}{=} \left(\lambda^4 k f_\infty\right)^t \alpha'_0 + (\lambda^2 m + n_w f_\infty + \lambda^4 mk + n_w)\sum_{i=0}^{t-1}\left(\lambda^4 k f_\infty\right)^i,
\end{aligned} \tag{43}$$

where (a) follows from (42) and (b) follows by recursively apply (a).

We observe from (43) that if $(\lambda^4 k f_\infty) = \left(\frac{\lambda^4 k \tilde{N}(\beta, P_R)}{(k_2+\sqrt{k_1 k})^2+\tilde{N}(\beta,P_R)}\right) < 1$, then the sequence $\{\alpha'_t\}$ converges to a limit point as $t \to \infty$ and consequently the original sequence $\{\alpha_t\}$ is guaranteed to stay bounded. Thus the system in (70) can be mean square stabilized over the given *half-duplex sensor network* if

$$\lambda^4 < \left(\frac{(k_2+\sqrt{k_1 k})^2 + \tilde{N}(\beta, \{P_R^i\}_{i=1}^L)}{k\tilde{N}(\beta, \{P_R^i\}_{i=1}^L)}\right) \tag{44}$$

$$\Rightarrow \log(\lambda) < \frac{1}{4}\log\left(\log\left(\frac{1}{k}\right) + \log\left(1+\frac{(k_2+\sqrt{k_1 k})^2}{\tilde{N}(\beta, P_R)}\right)\right)$$

$$= \log(\lambda) < \frac{1}{4}\left(\log\left(1+\frac{2h^2\beta P_S}{N_d}\right) + \log\left(1+\frac{\tilde{M}(\beta, \{P_R^i\}_{i=1}^L)}{\tilde{N}(\beta, \{P_R^i\}_{i=1}^L)}\right)\right), \tag{45}$$

where in the last equality we substituted $k = \frac{N}{2h^2\beta P_S + N}$ and $M(\beta, \{P_R^i\}_{i=1}^L) = (k_2+\sqrt{k_1 k})^2$ in order to show the dependencies on the average transmit powers of the L sensor nodes $\{P_R^i\}_{i=1}^L$ and the power allocation parameter β at the state encoder. Since the sensor nodes amplify the desired signal as well as the noise which is then superimposed at the decoder to the signal coming directly from the state encoder, an optimal choice of the sensor transmit power $\{P_R^i\}_{i=1}^L : \sum_{i=1}^L P_R^i \le P_R$ depends on the relay channel parameters $\{P_S, \{N_R^i\}_{i=1}^L, N_d, h, h_i, \beta\}$. Moreover an optimal choice of the power allocation factor β at the encoder also depends on the channel parameters $\{P_S, \{P_R^i\}_{i=1}^L, \{N_R^i\}_{i=1}^L, N, h, \{h_i\}_{i=1}^L\}$. Therefore we rewrite (45) as

$$\log(\lambda) <$$
$$\frac{1}{4}\max_{\substack{0<\beta\le 1\\ P_R^i:\sum_i P_R^i \le P_R}}\left\{\log\left(1+\frac{2h^2\beta P_S}{N_d}\right) + \log\left(1+\frac{\tilde{M}\left(\beta, \{P_R^i\}_{i=1}^L\right)}{\tilde{N}\left(\beta, \{P_R^i\}_{i=1}^L\right)}\right)\right\}, \tag{46}$$

which completes the proof of Theorem 5. ∎

Remark 5. *An optimal choice of the power allocation parameter β at the state encoder and and an optimal power allocation at the sensor nodes $\{P_R^i\}_{i=1}^L$ which maximize the term on the right hand side of (32) depend on the quality (i.e., signal-to-noise ratio) of $\mathcal{E}-\mathcal{D}$, $\mathcal{E}-\mathcal{R}_i$, and $\mathcal{R}_i-\mathcal{D}$ links.*

Remark 6. *The term on the right hand side of (32) is the information rate over the half-duplex AWGN relay channel with noiseless feedback. This is shown in [9, Appendix A]. For channels with feedback, directed information is a useful* *quantity [26]. We show in Appendix **??** that that the directed information rate over the given non-orthogonal half-duplex sensor network is also equal to the term on the right hand side of (32).*

Remark 7. *It is interesting to see that the sufficient condition for mean square stability in (32) does not depend on the process noise $\{W_t\}$. This provides motivation to study stabilizability of the system in* (70) *without process noise. Therefore in the following we analyze stability of a noiseless system under our proposed scheme.*

Noiseless Plant In the absence of the process noise in (70), the state variance of the noiseless system under our proposed scheme can be obtained by substituting $n_w = 0$ in (40) and (41). That is the state variance of the noiseless plant is given by

$$\alpha_t = \left(\frac{\lambda^2 N}{2h^2\beta P_S + N}\right)\alpha_{t-1}, \qquad t = 2, 4, 6, ...$$
$$\alpha_t = \left(\frac{\lambda^4 k\tilde{N}(\beta)}{(k_2 + \sqrt{k_1 k})^2 + \tilde{N}(\beta)}\right)\alpha_{t-2}, \qquad t = 3, 5, 7, ... \tag{47}$$

Since $\alpha_1 = \frac{\lambda^2 N}{h^2 P_S}\alpha_0 + n_w$, the state variance $\alpha_t \to 0$ as $t \to \infty$ if $\left(\frac{\lambda^4 k\tilde{N}(\beta)}{(k_2+\sqrt{k_1 k})^2+\tilde{N}(\beta)}\right) < 1$. This is the same condition as in (44). Thus by using the proposed linear coding and control scheme, we obtain identical sufficient conditions for mean square stability of noisy and noiseless first order LTI system over the *non-orthogonal half-duplex* sensor network. Although the sufficient conditions are identical, the state variance in the noisy plant scenario cannot converge to zero unlike the noiseless scenario.

A comparison of second moments of the plant state process at three different power levels of process noise is illustrated in Fig. 4. We fix the relay channel parameters $\{P_S = 1, P_R = 1, h = 1, \beta = 0.5, N = 0.5, N_R = 0.1\}$, the plant parameters $\{\alpha_0 = 0.25, \lambda = 1.5\}$, and plot second moment of the state process $\mathbb{E}[X_t^2]$ moment as a function of time t for three power levels of process noise, i.e., n_w = 0, 0.1, and 0.25. For the given set of channel parameters, *mean square stability* of the system requires $\lambda < 1.975$ according to Theorem 5. In Fig. 4 we have fixed $\lambda = 1.5$ (i.e., less than 1.975), therefore the second moment stays bounded for all levels of process noise. For $n_w = 0$ the second moment converges to zero, starting from an arbitrary an value equal to 0.25 as shown in Fig. 4. For non-zero values of process, the second moment keeps alternating between two different values. This happens due to the first and the second transmission phases in the *half-duplex* relay channel introduced in Sec. **??**. As shown in Fig. 4, for $n_w = 0.1$ and $n_w = 0.25$ the second moment converges to a unique non-zero value for each transmission phase and thus it keeps alternating between these two unique limit points. This will become more clear in Sec. **??**. In Fig. 4 we can also observe that the rate of convergence is similar for the three examples, and seems to be unaffected by the power level of process noise.

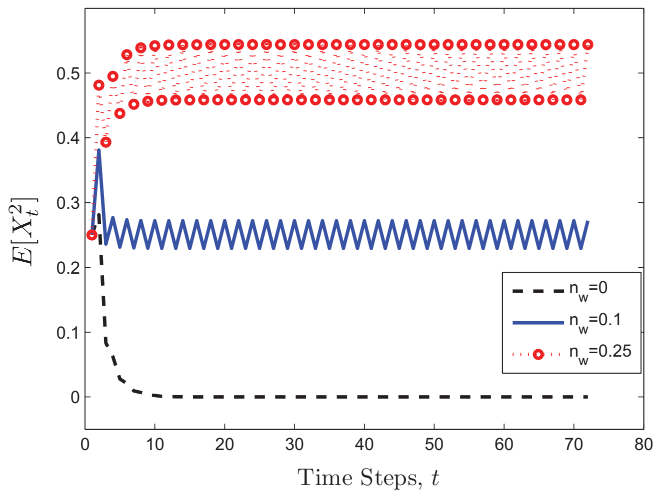


Figure 4: Comparison of second moments of the plant state process at three different levels of process noise.

Proposition 2. *For the noiseless plant, the following infinite horizon quadratic cost can be achieved:*

$$\sum_{t=1}^{\infty} \mathbb{E}\left[X_t^2 + qU_t^2\right] = \alpha_1 \frac{1 + \lambda^2 b_1 + q(\lambda^4 b_1 c_2 + \lambda^2 b_2)}{1 - \lambda^4 b_1 c_1}, \tag{48}$$

where $q > 0$, $\alpha_1 = \frac{\lambda^2 N}{h^2 P_S}\alpha_0$, $b_1 = \frac{N}{2h^2\beta P_S + N}$, $b_2 = \frac{2h^2\beta P_S}{2h^2\beta P_S + N}$, $c_1 = \frac{\tilde{N}(\beta,P_R)}{\tilde{M}(\beta,P_R)+\tilde{N}(\beta,P_R)}$, *and* $c_2 = \frac{\tilde{M}(\beta,P_R)}{\tilde{M}(\beta,P_R)+\tilde{N}(\beta,P_R)}$.

Proof. This has been shown in [9, Appendix B]. ■

By choosing certain values of the parameters β and h, we can get special cases of the general half-duplex sensor network. One special case has been discussed below.

Two-Hop Sensor Network We now consider a special case of the half-duplex sensor (relay) network illustrated in Fig. 3.3 with $h = 0$. This special case models a realistic network scenario where the state information can be communicated to the remote controller only via the sensor nodes and there is no way for the state encoder to directly communicate with the controller. We call this network as a *two-hop* sensor network, where the communication from the state encoder to the controller takes place in two hops. In the first hop the sensor nodes receive the state information from the state encoder, which then relay the state information to the controller in the second hop. We can obtain a sufficient condition for the mean square over this network by substituting $h = 0, \beta = 1$ in Theorem 5.

Corollary 3. *The scalar linear time invariant system in (70) can be mean square stabilized over the* two-hop non-orthogonal half-duplex *sensor network if*

$$\log(\lambda) < \frac{1}{4} \max_{P_R^i : \sum_i P_R^i \le P_R} \left\{ \log \left(1 + 2P_S \frac{\left(\sum_{i=1}^L \sqrt{\frac{h_i^2 P_R^i}{2P_S + N_R^i}} \right)^2}{\sum_{i=1}^L \frac{h_i^2 P_R^i N_R^i}{2P_S + N_R^i} + N_d} \right) \right\}. \tag{49}$$

Theorem 6. *The scalar linear time invariant system in (70) cannot be mean square stabilized over the given* two non-orthogonal half-duplex *sensor network if*

$$\log(\lambda) \ge \frac{1}{2} \log \frac{N_R + LP_S}{N_R + LP_S 2^{-2R}}, \tag{50}$$

where

$$R = \frac{1}{2} \log \left(1 + \frac{1}{N_d} \left(\sum_{i=1}^L h_i^2 P_R^i + 2 \sum_{i=1}^L \sum_{k=i+1}^L \rho_{i,k}^\star h_i h_k \sqrt{P_R^i P_R^k} \right) \right), \tag{51}$$

and $N_R = \frac{L}{\sum_{i=1}^L (1/N_r^i)}$, *and* $\rho_{i,k}^\star = \frac{P_S}{(P_S + N_R)}$.

Proof. The proof can be found in [41]. ∎

Theorem 7. *The scalar linear time invariant system in (70) can be mean square stabilized over the a symmetric* two-hop half-duplex *sensor network with a linear state encoder and* $\{P_R^i = \frac{P_R}{L}, h_i = 1, N_R^i = N_R\}$ *for all* $i \in \{1, 2, \ldots, L\}$, *if and only if*

$$\log(\lambda) < \frac{1}{4} \log \left(1 + \frac{2LP_S P_R}{2P_S N_d + P_R N_R + N_R N_d} \right). \tag{52}$$

Proof. The proof can be found in [41]. ∎

3.4 Non-orthogonal Full-duplex Sensor Network

We now consider a network of full-duplex sensors nodes which are capable of transmitting and receiving signals simultaneously unlike the half-duplex scenario discussed in Sec. 3.3. Further we assume that the state encoder and all the sensor nodes transmit signals in the same time slot and the same frequency band, that is we have non-orthogonal signal transmissions in the network. This *non-orthogonal full-duplex* sensor network scenario is illustrated in 5. At any discrete time step t the state encoder $\mathcal{E}$ transmits $S_{e,t}$ with an average power P_S. The sensor nodes $\{\mathcal{R}_i\}_{i=1}^{L}$ receive $S_{e,t}$ as,

$$Y_t^i = S_{e,t} + Z_{r,t}^i, \qquad i \in \{1, 2, \ldots, L\}, \tag{53}$$

where $Z_{r,t}^i \sim \mathcal{N}(0, N_r^i)$ are the additive white Gaussian noise components which are independent both across time and across the sensor nodes. The sensor node i transmits $S_{r,t}^i$ with an average power P_R^i such that $\sum_{i=1}^{L} P_R^i \leq P_R$. Since the sensors are full-duplex, the decoder $\mathcal{D}$ at the control unit receives,

$$R_t = hS_{e,t} + \sum_{i=1}^{L} h_i S_{r,t}^i + Z_{d,t}, \tag{54}$$

where $Z_{d,t} \sim \mathcal{N}(0, N_d)$ is a white Gaussian noise sequence. The variables $h, h_i \in \mathbb{R}$ denote the gains of $\mathcal{E} - \mathcal{D}$ link and $\mathcal{R}_i - \mathcal{D}$ link for all $i \in \{1, 2, \ldots, L\}$. In the following we present a sufficient condition for mean square stability of the first order linear system in (70) over the given *non-orthogonal full-duplex* sensor network.

Theorem 8. *The linear scalar LTI system in (70) with $W_t = 0$ can be mean square stabilized over the* non-orthogonal full-duplex *Gaussian sensor network if*

$$\log(\lambda) < \frac{1}{2} \max_{P_R^i : \sum_{i=1}^{L} P_R^i \leq P_R} \left\{ \log \left(1 + \frac{\left(\sqrt{h^2 P_S} + \eta^\star \sum_{i=1}^{L} \sqrt{\frac{h_i^2 P_S P_R^i}{P_S + N_R^i}} \right)^2}{N_d + \sum_{i=1}^{L} \frac{h_i^2 P_R^i N_R^i}{P_S + N_R^i}} \right) \right\}, \tag{55}$$

where $\eta^\star$ is the unique root in the interval $[0, 1]$ of the following fourth order

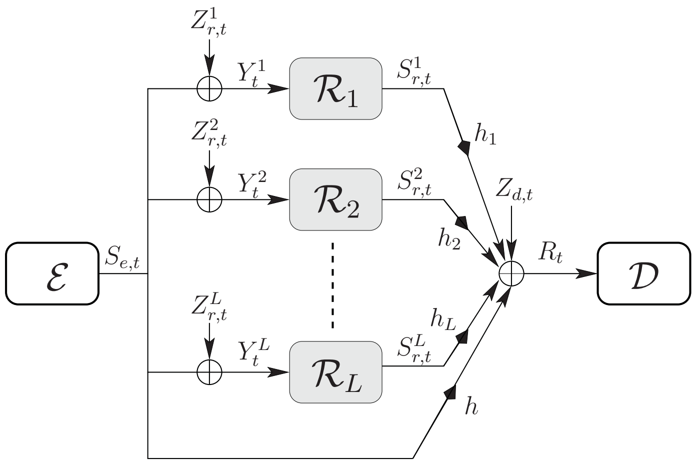


Figure 5: The unstable plant has to be controlled by the actions of observer/encoder ($\mathcal{O}/\mathcal{E}$) and decoder/controller ($\mathcal{D}/\mathcal{C}$) over the AWGN relay channel.

polynomial

$$\left(\sum_{i=1}^{L}\sqrt{\frac{h_i^2 P_S P_R^i}{(P_S+N_R^i)}}\right)\eta^4+\left(2hP_S\sum_{i=1}^{L}\sqrt{\frac{h_i^2 P_R^i}{(P_S+N_R^i)}}\right)\eta^3$$
$$+\left(h^2P_S+N_d+\sum_{i=1}^{L}\frac{h_i^2P_R^iN_R^i}{P_S+N_R^i}\right)\eta^2=\left(N_d+\sum_{i=1}^{L}\frac{h_i^2P_R^iN_R^i}{P_S+N_R^i}\right).$$

Remark 8. *The term on the right hand side of the inequality in (69) is the achievable rate over the* non-orthogonal full-duplex *AWGN relay channel [62, Theorem 5].*

Proof. In order to prove Theorem 8 we propose to use a linear and memoryless communication and control scheme. This scheme is in principal similar to the scheme we proposed for the *half-duplex* sensor network in Sec. 3.3 with some modifications to adapt to full-duplex nature of the sensor nodes. A full-duplex sensor node can simultaneously transmit and receive signals, therefore in this scheme the sensor nodes transmit in every time slot in contrast to the the half-duplex network scenario where the sensor nodes transmit in alternate time slots. The initial transmission and control at $t = 0$ in the *full-duplex* scenario is identical to that of the scheme proposed for the *half-duplex* scenario in Sec. 3.3. Therefore according to Sec. 3.3, the plant state $X_1 \sim \mathcal{N}(0, \alpha_1)$ where $\alpha_1 = \frac{\lambda^2 N_d}{h^2 P_S}\alpha_0 + n_w$. Further transmissions and control actions work as follows.

Time step $t = 1$

The encoder $\mathcal{E}$ observes X_1 and it then transmits $S_{e,1} = \sqrt{\frac{P_S}{\alpha_1}} X_1$ to the decoder $\mathcal{D}$ at the remote control unit. The sensor nodes $\{\mathcal{R}_i\}_{i=1}^{L}$ overhear the signal transmitted by the controller but remain silent. The decoder $\mathcal{D}$ observes $R_1 = hS_{e,1} + Z_1$ and computes the MMSE estimate of X_1, which is given by

$$\hat{X}_1 = \mathbb{E}[X_1|R_1] \overset{(a)}{=} \frac{\mathbb{E}[X_1 R_1]}{\mathbb{E}[R_1^2]} R_1 \overset{(b)}{=} \left(\frac{h\sqrt{P_S \alpha_1}}{h^2 P_S + N_d}\right) R_1,$$

where (a) follows from the fact that the optimum MMSE estimator for a Gaussian variable is linear [34]; and (c) follows from $\mathbb{E}[X_1 R_1] = \sqrt{h^2 P_S \alpha_1}$ and $\mathbb{E}[R_1^2] = h^2 P_S + N$.

The controller $\mathcal{C}$ takes an action $U_1 = -\lambda \hat{X}_1$ which results in $X_2 = \lambda(X_1 - \hat{X}_1) + W_1$. The new plant state X_2 is linear combination of zero mean Gaussian variables $\{X_1, \hat{X}_1, W_1\}$, therefore it is also zero mean Gaussian with the following variance

$$\begin{aligned} \alpha_2 \triangleq \mathbb{E}[X_2^2] &= \lambda^2 \mathbb{E}[(X_1 - \hat{X}_1)^2] + \mathbb{E}[W_1^2] \\ &= \lambda^2 \left(\frac{N_d}{h^2 P_S + N_d}\right) \alpha_1 + n_w, \end{aligned} \tag{56}$$

where the last equality follows from $\mathbb{E}[X_1 \hat{X}_1] = \mathbb{E}[\hat{X}_1^2] = \frac{h^2 P_S \alpha_1}{h^2 P_S + N_d}$ (by computation).

Time steps $t \geq 2$

The encoder $\mathcal{E}$ observes X_t and it then transmits $S_{e,t} = \sqrt{\frac{P_S}{\alpha_t}} X_t$. The sensor nodes simultaneously receive this signal and transmit an amplified version of the signal they had received in the previous time step under an average power constraint, i.e., the sensors merely act as amplify-and-forward relays. Thus the signal transmitted by the i-th sensor is given by

$$S_{r,t}^i = \sqrt{\frac{P_R^i}{P_S + N_R^i}} \left(S_{e,t-1} + Z_{r,t-1}^i\right), \qquad \text{for all } i \in \{1, 2, \cdots, L\}, \tag{57}$$

where $\mathbb{E}[(S^i_{r,t})^2] = P^i_R$ and the signal amplification is done to ensure $\sum_{i=1}^L P^i_R \leq P_R$. Accordingly the decoder $\mathcal{D}$ receives

$$R_t = hS_{e,t} + \sum_{i=1}^{L} h_i S_{r,t} + Z_t = L_1 X_t + L_2 X_{t-1} + \tilde{Z}_t, \tag{58}$$

where $L_1 = \sqrt{\frac{h^2 P_S}{\alpha_t}}$, $L_2 = \sum_{i=1}^L \sqrt{\frac{h_i^2 P_S P_R^i}{(P_S+N_R^i)\alpha_{t-1}}}$, and $\tilde{Z}_t = Z_t + \sum_{i=1}^L \sqrt{\frac{h_i^2 P_R^i}{P_S+N_R^i}} Z^i_{r,t-1}$ with $\tilde{Z}_t \sim \mathcal{N}(0, \tilde{N})$. The computation of the state MMSE estimate $\hat{X}_t$ and the action taken by controller $U_t = -\lambda \hat{X}_t$ are identical to that of the *half-duplex* sensor network scheme proposed in Sec. 3.3. Therefore according to (38) the variance α_{t+1} of the new plant state X_{t+1} (obtained after taking the control action U_t) is given by

$$\begin{aligned}
\alpha_{t+1} &\triangleq \mathbb{E}[X^2_{t+1}] = \lambda^2 \alpha_t \left(\frac{L_2^2 n_w + \lambda^2 \tilde{N}(P_R)}{(\lambda L_1 + L_2)^2 \alpha_t + L_2^2 n_w + \lambda^2 \tilde{N}(P_R)} \right) + n_w \\
&\overset{(a)}{=} \lambda^2 \alpha_t \left(\frac{\left(\frac{n_w k_1}{\lambda^2}\right) \frac{1}{\alpha_{t-1}} + \tilde{N}(P_R)}{\left(k_2 + \sqrt{\frac{k_1}{\lambda^2}\frac{\alpha_t}{\alpha_{t-1}}}\right)^2 + \left(\frac{n_w k_1}{\lambda^2}\right)\frac{1}{\alpha_{t-1}} + \tilde{N}(P_R)} \right) + n_w,
\end{aligned} \tag{59}$$

where (a) follows by substituting the values of L_1 and L_2; and by defining $k_1 = \left(\sum_{i=1}^L \sqrt{\frac{h_i^2 P_S P_R^i}{(P_S+N_R^i)}}\right)^2$ and $k_2 = \sqrt{h^2 P_S}$.

Our aim is to find condition on the system parameter λ which is sufficient to ensure that the state variance in (59) stays bounded. In order to simplify the problem we assume that there is no process noise in the system. By substituting $n_w = 0$ in (59) we get

$$\alpha_{t+1} = \left(\frac{\lambda^2 \tilde{N}(P_R)}{\left(k_2 + \sqrt{\frac{k_1}{\lambda^2}\frac{\alpha_t}{\alpha_{t-1}}}\right)^2 + \tilde{N}(P_R)} \right) \alpha_t, \qquad \forall t \geq 2. \tag{60}$$

By defining $\eta_t \triangleq \sqrt{\frac{1}{\lambda^2}\frac{\alpha_t}{\alpha_{t-1}}}$, we can rewrite (60) as

$$\eta_{t+1} = \sqrt{\frac{\tilde{N}(P_R)}{\left(k_2 + \eta_t \sqrt{k_1}\right)^2 + \tilde{N}(P_R)}}, \qquad \forall t \geq 2. \tag{61}$$

We will now show that the sequence $\{\eta_t\}$ converges to a unique fixed point. The convergence follows from the following lemma.

Lemma 2. *([62, Lemma 2]) Consider the function $f : x \mapsto \sqrt{\frac{a}{a+p(1+bx)^2}}$ defined on the closed interval [0, 1] when $a, b, p \geq 0$. The function $f(.)$ has exactly one fixed point $x^* \in [0,1]$ and the infinite sequence $x_0, x_1 = f(x_0), x_2 = f(x_1), \ldots$ converges to this fixed point for any starting point $x_0 \in [0,1]$.*

According to Lemma 2, starting with $\eta_2 = \frac{\alpha_2}{\alpha_1} = \sqrt{\frac{N}{h^2 P_S + N}} \in [0,1]$, the sequence $\{\eta_t\}$ in (61) converges to a fixed point $\eta^\star$. This fixed point is given by the unique solution in the interval [0,1] of the following fourth order polynomial

$$\left(\sum_{i=1}^{L}\sqrt{\frac{h_i^2 P_S P_R^i}{(P_S+N_R^i)}}\right)\eta^4 + \left(2hP_S\sum_{i=1}^{L}\sqrt{\frac{h_i^2 P_R^i}{(P_S+N_R^i)}}\right)\eta^3$$
$$+\left(h^2P_S + N_d + \sum_{i=1}^{L}\frac{h_i^2 P_R^i N_R^i}{P_S+N_R^i}\right)\eta^2 = \left(N_d + \sum_{i=1}^{L}\frac{h_i^2 P_R^i N_R^i}{P_S+N_R^i}\right),$$

where the above polynomial has been obtained by simplifying (61) and substituting the values of $k_1, k_2, \tilde{N}(P_R)$ [62].

Having shown that the sequence $\{\eta_t\}$ converges to the unique fixed point, we now find the values of the system parameter λ for which the state variance $\{\alpha_t\}$ converges to a limit point and consequently stays bounded. Rewriting (60) as

$$\begin{aligned}\alpha_{t+1} &= \left(\frac{\lambda^2\tilde{N}(P_R)}{\left(k_2+\eta_t\sqrt{k_1}\right)^2+\tilde{N}(P_R)}\right)\alpha_t \\ &= \left(\prod_{i=2}^{t}\left(\frac{\tilde{N}(P_R)}{\left(k_2+\eta_i\sqrt{k_1}\right)^2+\tilde{N}(P_R)}\right)\right)\lambda^{2t-2}\alpha_2 \\ &= \left(\mu(t)\lambda^{2t}\right)\nu = \left(\lambda^{t\left[\frac{1}{t}\log_\lambda(\mu(t))+2\right]}\right)\nu,\end{aligned} \tag{62}$$

where $\mu(t) \triangleq \prod_{i=2}^{t}\left(\frac{\tilde{N}(P_R)}{\left(k_2+\eta_i\sqrt{k_1}\right)^2+\tilde{N}(P_R)}\right)$ and $\nu \triangleq \lambda^{-2}\alpha_2$.

We observe from (62) that $\alpha_t \to 0$ as $t \to \infty$ if

$$\lim_{t\to\infty}\left[\frac{1}{t}\log_\lambda\left(\mu\left(t\right)\right)+2\right]<0, \tag{63}$$

where the existence of the limit follows from convergence of the sequence $\{\eta_i\}$.

Since $\log_\lambda(\mu(t)) = \frac{\log(\mu(t))}{\log(\lambda)}$, we can rewrite (63) as

$$\log(\lambda) < \lim_{t\to\infty} \frac{1}{2t}\log\left(\frac{1}{\mu(t)}\right) \overset{(a)}{=} \lim_{t\to\infty}\frac{1}{2t}\log\left(\prod_{i=2}^{t}\left(1+\frac{\left(k_2+\eta_i\sqrt{k_1}\right)^2}{\tilde{N}(P_R)}\right)\right)$$
$$= \lim_{t\to\infty}\frac{1}{2t}\sum_{i=2}^{t}\log\left(1+\frac{\left(k_2+\eta_i\sqrt{k_1}\right)^2}{\tilde{N}(P_R)}\right) \overset{(b)}{=} \frac{1}{2}\log\left(1+\frac{\left(k_2+\eta^\star\sqrt{k_1}\right)^2}{\tilde{N}(P_R)}\right)$$
$$\overset{(c)}{=} \frac{1}{2}\log\left(1+\frac{\left(\sqrt{h^2P_S}+\eta^\star\sum_{i=1}^{L}\sqrt{\frac{h_i^2P_SP_R^i}{P_S+N_R^i}}\right)^2}{N_d+\sum_{i=1}^{L}\frac{h_i^2P_R^iN_R^i}{P_S+N_R^i}}\right),$$

where (a) follows by substituting the value of $\mu(t)$ from (62); (b) follows by convergence of the sequence $\{\eta_i\}$ to $\eta^\star$ and using Cesaro mean theorem [**?**]; and (c) follows by substituting the values of k_1, k_2, and $\tilde{N}(P_R)$. This completes the proof of Theorem 8. ∎

3.5 Cascade Sensor Network

A cascade sensor network is depicted in Fig. 6, where there are L sensor nodes $\{\mathcal{R}_i\}_{i=1}^{L}$ connected in series. The communication channels between all the nodes are modeled as white Gaussian channels, that is in Fig. 6 the variables $Z_t^i \sim \mathcal{N}(0, N_i)$ denote mutually independent white Gaussian noise components for $i = \{1, 2, \cdots, L+1\}$. In the given cascade sensor network setup, the state encoder $\mathcal{E}$ observes state of the system and transmits $S_{e,t}$ with an average power P_S, which is received by the first sensor node as $Y_t^1 = S_{e,t} + Z_t^1$. Upon receiving the signal from the encoder, the first sensor node transmits the state information to the second sensor node and so on. That is at any time step t, the sensor node $\mathcal{R}_i$ receives

$$Y_t^i = S_{r,t}^{i-1} + Z_t^i, \tag{64}$$

and it then transmits $S_{r,t}^i$ with an average power P_R^i to the next sensor node $\mathcal{R}_{i+1}$ for all $i \in \{2, 3, \cdots, L\}$. There is a total average transmit power constraint on the sensor nodes, $\sum_{i=1}^{L} P_R^i \leq P_R$. Finally the decoder $\mathcal{D}$ at the control unit receives $Y_t^{L+1} = S_{r,t}^L + Z_t^{L+1}$, which then takes control action to stabilize the system.

For mean square stabilizing the first order LTI system in (70) over the given cascade sensor network we can employ a linear memoryless Schalkwijk-Kailath based scheme, like we did for the network scenarios discussed earlier. By a similar analysis we obtain the following sufficient condition for mean square stability of the system.

Theorem 9. *The scalar linear time invariant system in (70) can be mean square stabilized over a* cascade sensor network *of L relay nodes if*

$$\log(\lambda) < \frac{1}{2L}\log\left(1+\frac{P_S}{P_S+N_1}\prod_{i=1}^{L}\left(\frac{P_R^i}{P_R^i+N_{i+1}}\right)\right), \tag{65}$$

where the optimal choice of power allocation is $P_R^i = \frac{P_R}{L}$.

Proof. The proof can be found in [41]. ■

The above sufficient condition has been obtained using linear policies at the state encoder and at the sensor nodes. However from [7, 28] we know that the linear policies are not optimal for transmission of a Gaussian source over a Gaussian cascade sensor network. In [28], the authors studied a multi-stage decision (encoding) problem and provided counter-examples (based on functions whose output can take on only two values) to show that linear policies are not optimal when the number of stages are sufficiently large. According to their result linear encoding/sensing policies are not optimal when the number of sensor nodes in our model of cascade network is greater than two. We extended the work of [28] in [7] by studying a setup when there is only a single relay (sensor) node between the source node and the destination node. We show that although linear encoding policies are person-by-person optimal for this simple setup, they are not globally optimal in general.

Theorem 10. *The scalar linear time invariant system in (70) can not be mean square stabilized over the* cascade sensor network *if*

$$\log(\lambda) \geq \frac{1}{2L}\min\left\{\log\left(1+\frac{P_S}{N_1}\right), \log\left(1+\frac{P_R^1}{N_2}\right), \ldots, \log\left(1+\frac{P_R^L}{N_{L+1}}\right)\right\}. \tag{66}$$

Proof. The proof follows from information theoretic cut-set bound. ■

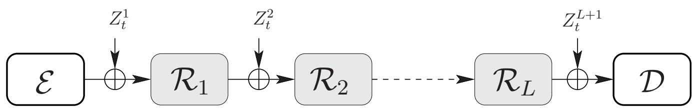


Figure 6: Control over a cascade sensor network.

3.6 Parallel Sensor Network

Consider the sensor network shown in Fig. 7, where the signal transmitted by a sensor node does not interfere with the signals transmitted by the other sensor nodes, i.e., there are L parallel channels from $\{\mathcal{R}_i\}_{i=1}^L$ to $\mathcal{D}$. We name this setup as a *parallel sensor network*, which models a practical scenario where the signal spaces of the sensor nodes are mutually orthogonal. For example the signals may be transmitted in either disjoint frequency bands or in disjoint time slots. According to Fig. 7, the signal received by each sensor node is given by

$$Y_t^i = S_{e,t} + Z_{r,t}^i, \qquad \text{for all } i \in \{1, 2, \dots, L\}, \tag{67}$$

where $S_{e,t}$ is the signal transmitted by the state encoder $\mathcal{E}$ with an average power P_S and $Z_{r,t}^i \sim \mathcal{N}(0, N_r^i)$ is a Gaussian noise sequence, which is independent both across time and across the sensor nodes. Upon receiving Y_t^i the sensor node $\mathcal{R}_i$ transmits $S_{r,t}^i$ subject to an average power constraint P_R^i. Accordingly the decoder receives,

$$R_t^i = S_{r,t}^i + Z_{d,t}^i, \qquad \text{for all } i \in \{1, 2, \dots, L\}, \tag{68}$$

where $Z_{d,t}^i \sim \mathcal{N}(0, N_d^i)$ is a Gaussian noise sequence, which is independent both across time and across the sensor nodes. In the following we present a sufficient condition for mean square stability of the system in (70) over the given sensor network.

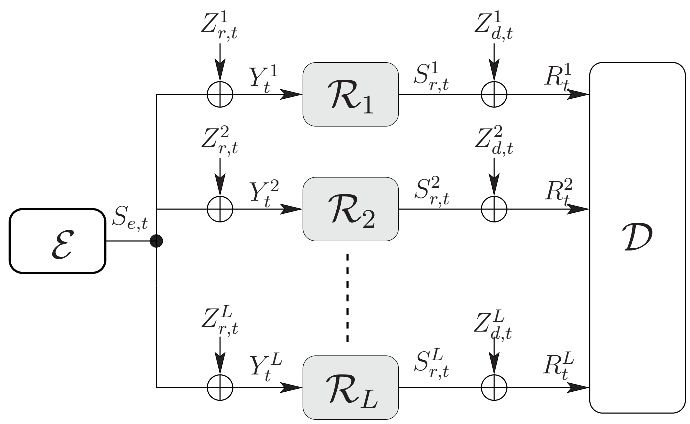


Figure 7: Control over a parallel sensor network.

Theorem 11. *The linear scalar system in (70) cannot be mean square stabilized over the Gaussian* parallel sensor network *if*

$$\log(\lambda) \geq \frac{1}{2} \min \left\{ \log\left(1 + \sum_{i=1}^{L} \frac{P_S}{N_R^i}\right), \max_{P_R^i : P_R \geq 0, \sum_i P_R^i \leq P_R} \sum_{i=1}^{L} \log\left(1 + \frac{P_R^i}{N_d^i}\right) \right\} \tag{69}$$

Proof. The proof follows from information theoretic cut-set arguments. ∎

For the *parallel sensor network* we can derive sufficient condition for mean square stability using linear policies policies like previously discussed scenarios. However we know that linear policies are highly sub-optimal for parallel sensor network setting. A distributed joint source-channel code is optimal in minimizing mean-square distortion if the following two conditions hold: i) All channels from the source to the destination send independent information. ii) All channels utilize the capacity, i.e., source-channel needs to be matched. If we use linear linear policies at the sensors then the first condition is not fulfilled because all sensors would be transmitting correlated information. In [130] the authors proposed a non-linear scheme for a parallel network of two sensors, in which one sensor transmits only magnitude of observed state and the other sensor transmits only phase value (plus or minus sign) of the observed state. The magnitude and phase of the state were shown to be independent and thus the scheme fulfilled the first condition of optimality. This nonlinear sensing scheme was shown to outperform the best linear scheme for the LQG control problem, although the second condition of source-channel matching is not fulfilled by this non-linear scheme. We can use this non-linear scheme together with an SK type scheme which will ensure source-channel matching by making the outputs of the two sensors Gaussian distributed after the initial transmissions as shown earlier in Sections 3.3 and 3.4.

3.7 Conclusion

We studied the problem of mean square stabilizing a discrete time LTI system over some basic topologies of Gaussian sensor networks. We proposed to use delay-free linear communication and control strategies, and thereby obtained sufficient conditions for stabilization. We also obtained necessary conditions for stabilization using information theoretic bounds and in some cases bounds are shown to be tight. Our results reveal a relationship between the communication channel parameters (i.e., signal-to-noise ratios) and the possibility of stabilizing the plant. Some discussion on vector valued systems can be found in [41]. An interesting extension of this work would be to consider instantaneous non-linear relaying strategies which can potentially increase the achievable rate and thus extend the class of stablizable systems over the considered Gaussian sensor networks.

4 Closed-loop Control Over Basic Multi-user Communication Channels

In this section we consider the problem of remotely controlling scalar linear time invariant systems over basic multi-user communication channels. We study three basic communications channels for stabilizing two LTI systems: i) white Gaussian multiple-access channel, ii) white Gaussian broadcast channel, and iii) white Gaussian interference channel.

A two user multiple-access channel is the communication channel where two sources transmit their messages to a common destination [46]. By the control over the multiple-access channel we mean that there exist two separate sensors to sense the states and a single remote controller to stabilize the two plants i.e., a multi-sensor joint controller setup.The capacity region of the two-user memoryless Gaussian multiple-access channel with noiseless feedback is found in [22], which is relevant to the problem of closed-loop control over the multiple-access channel.

A two user broadcast channel is the communication channel where one sender transmits messages to two destinations [46]. Control over broadcast channel refers to a joint sensor multi-controller setup i.e., there exists a common sensor to jointly observe the states of the two plants and there are two separate remote controllers in order to stabilize them. The capacity region of the Gaussian broadcast channel with and without feedback is not known [46]. For the problem of closed-loop control, the broadcast channel with feedback is more relevant. In [23] Ozarow et al. provided an achievable rate region over the two user memoryless Gaussian broadcast channel with noiseless feedback which is highly relevant to our problem.

A two user interference channel is a fundamental communication channel where two sources wish to communicate their messages to two different destinations and the signals transmitted from the sources interfere with each other [6]. By the control over the interference we mean that there exist two separate sensors to sense the states of the two plants, and there exist two separate remote controllers to separately stabilize the two plants i.e., a multi-sensor multi-controller setup. The capacity of the general interference channel is still an open problem, however the capacity region is known for some special cases. However in the context of closed-loop control, the interference channel with feedback is more relevant. In [2, 5], the authors provided achievable rate regions over memoryless interference channel with noiseless and noisy feedback which is highly relevant to our problem. The coding schemes proposed by Kramer and Gastpar et al. in [2, 5] for the memoryless Gaussian interference channels with noiseless feedback are adaptations of the well-known Schalkwijk-Kailath coding scheme for memoryless Gaussian point-to-point communication channel with noiseless feedback [118].

The results presented in this section have appeared in [8, 12].

4.1 Problem Setup

We consider two scalar discrete-time LTI systems whose state equations are given by

$$\begin{aligned} X_{i,t+1} &= \lambda_i X_{i,t} + U_{i,t} + W_{i,t} \\ Y_{i,t} &= X_{i,t} + V_{i,t} \qquad \text{for} \quad i = 1, 2, \end{aligned} \tag{70}$$

where $\{X_{i,t}\} \subseteq \mathbb{R}$, $\{U_{i,t}\} \subseteq \mathbb{R}$, $\{Y_{i,t}\} \subseteq \mathbb{R}$, $\{W_{i,t}\} \subseteq \mathbb{R}$, and $\{V_{i,t}\} \subseteq \mathbb{R}$ are state, control, observation, process noise and measurement noise processes of the plant

i. We assume that the open-loop systems are unstable ($\lambda_i > 1$) and the initial states $X_{i,0}$ are random variables with arbitrary probability distributions having variance $\alpha_{i,0} = \mathbb{E}[X_{i,0}^2]$ and correlation coefficient $\rho_0 = \frac{\mathbb{E}[X_{1,0}X_{2,0}]}{\sqrt{\alpha_{1,0}\alpha_{2,0}}}$. We study the problem of remotely controlling the two unstable systems over the white Gaussian broadcast and multiple-access channels.

Control over multiple-access channel The setup for control over multiple-access channel is depicted in Fig. 8. There are separate observers $\mathcal{O}_1$ and $\mathcal{O}_2$ for the two plants, and there is a common control unit $\mathcal{C}$ situated at remote location. In order to communicate the observed state values to the controller, an encoder $\mathcal{E}_i$ is lumped with $\mathcal{O}_i$ and a decoder $\mathcal{D}$ is lumped with the controller. At any time instant t, the encoders $\mathcal{E}_1$ and $\mathcal{E}_2$ transmit $S_{1,t}$ and $S_{2,t}$ respectively, and the decoder $\mathcal{D}$ receives $R_t = S_{1,t} + S_{2,t} + Z_t$, where $Z_t \sim \mathcal{N}(0, N)$ is the white noise component. Let $f_{i,t}$ denote the observer/encoder policy for the plant i, then we have $S_{i,t} = f_{i,t}(\{Y_{i,k}\}_{k=0}^t)$ which must satisfy an average power constraint $\lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}[S_{i,t}^2] \leq P_i$. Further let $\gamma_{i,t}$ denote the decoder/controller policy, then $U_{i,t} = \gamma_{i,t}(\{R_k\}_{k=0}^t)$.

Control over broadcast channel The setup for control over broadcast channel is depicted in Fig. 9. There is a common observer $\mathcal{O}$ and separate controllers $\mathcal{C}_1$ and $\mathcal{C}_2$ for the two plants. In order to communicate the observed state values to the controllers, an encoder $\mathcal{E}$ is lumped with the observer and the decoders $\mathcal{D}_1$ and $\mathcal{D}_2$ are lumped with the respective controllers. At any time instant t, the encoder transmits S_t and the decoder $\mathcal{D}_i$ receives $R_{i,t} = S_t + Z_t + Z_{i,t}$, where $Z_{i,t} \sim \mathcal{N}(0, N_i)$ and $Z_t \sim \mathcal{N}(0, N)$ are the mutually independent white noise components. The noise component Z_t in the broadcast channel can model a common noise or interference in the two signals. Let f_t denote the observer/encoder policy, then we have $S_t = f_t(\{Y_{1,k}\}_{k=0}^t, \{Y_{2,k}\}_{k=0}^t)$ which must satisfy an average power constraint $\lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}[S_t^2] \leq P$. Further let $\gamma_{i,t}$ denote the decoder/controller policy, then $U_{i,t} = \gamma_{i,t}(\{R_{i,k}\}_{k=0}^t)$.

Control over Interference channel The setup for control over symmetric white Gaussian interference channel is depicted in Fig. 10. There are two separate observers $\{\mathcal{O}_1, \mathcal{O}_2\}$ and separate controllers $\{\mathcal{C}_1, \mathcal{C}_2\}$ for the two plants. In order to communicate the observed state values to the controllers, an encoder $\mathcal{E}_i$ is lumped with the observer $\mathcal{O}_i$ and a decoder $\mathcal{D}_i$ is lumped with the controller $\mathcal{C}_i$. At any

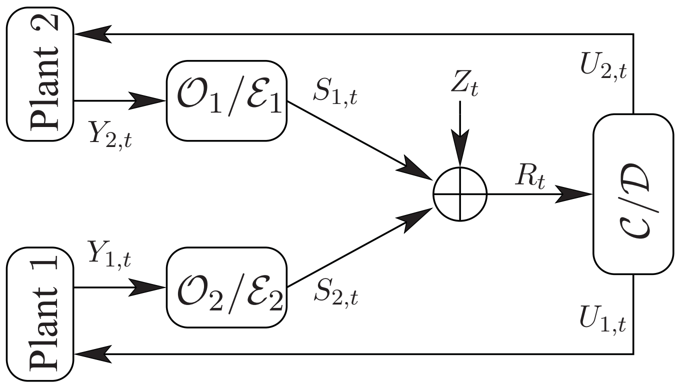


Figure 8: The two unstable LTI plants have to be controlled over the white Gaussian multiple access channel. There are two sensors to separately sense the states of the two plants and there is a remote common control unit.

time instant t, the encoders $\mathcal{E}_1$ and $\mathcal{E}_2$ transmit $S_{1,t}$ and $S_{2,t}$ respectively. Accordingly the decoder $\mathcal{D}_i$ receives $R_{i,t}$ which is given by

$$R_{1,t} = S_{1,t} + hS_{2,t} + Z_{1,t},$$
$$R_{2,t} = S_{2,t} + hS_{1,t} + Z_{2,t},$$

where $h \in \mathbb{R}_+$ is the cross channel gain, and $Z_{1,t} \sim \mathcal{N}(0, N)$ and $Z_{2,t} \sim \mathcal{N}(0, N)$ are white noise components with a fixed cross-correlation coefficient $\rho_z \triangleq \frac{\mathbb{E}[Z_{1,t}Z_{2,t}]}{N}$ in the interval $[-1, 1]$. The cross-correlation between the two noise components can model a common noise or common interference in the two signals. Let $f_{i,t}$ denote the ith observer/encoder policy, then we have $S_{i,t} = f_{i,t}(\{X_{i,k}\}_{k=0}^{t})$ which must satisfy an average power constraint $\lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}[S_{i,t}^2] \leq P$. Further let $\gamma_{i,t}$ denote the ith decoder/controller policy, then $U_{i,t} = \gamma_{i,t}(\{R_{i,k}\}_{k=0}^{t})$.

Mean square stability We assume that the process noise $W_{i,t}$ and the measurement noise $V_{i,t}$ in (70) are zero, and focus on mean square stability [1, 42, 108, 109, 116, 120] of the two plants. For a noise-free plant, we define mean square stability as follows.

Definition 2. *A system is said to be* mean square stable *if and only if*

$$\lim_{t\to\infty} \mathbb{E}[X_t^2] = 0,$$

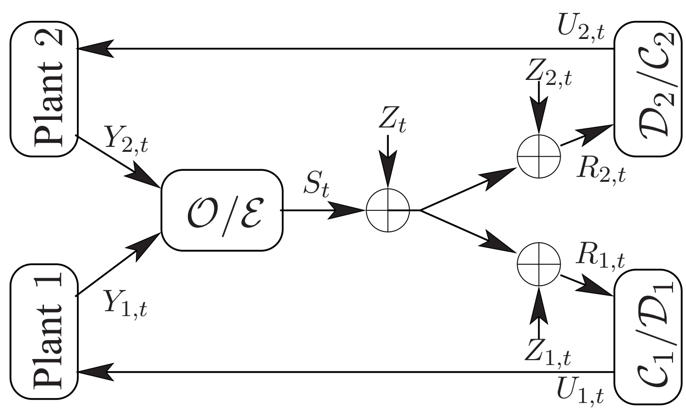


Figure 9: The two unstable LTI plants have to be controlled over the white Gaussian broadcast channel. There is a common sensor to jointly sense the states of the two plants and there are remotely located separate control units.

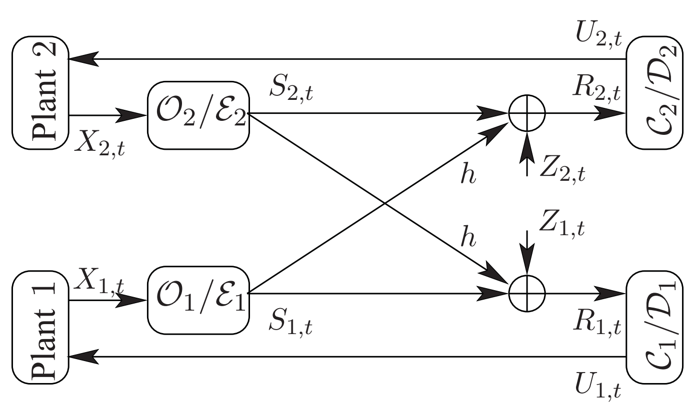


Figure 10: The two unstable LTI plants have to be controlled over the white Gaussian interference channel with correlated noise components. There are two sensors to separately sense the states and two remotely located control units to separately control the two plants.

regardless of the initial state X_0.

4.2 Main Results

We will first present our results in a comprehensive fashion and then provide the proofs in the next section.

Stability results for the multiple-access channel

Theorem 12. *The two scalar LTI systems in (70) with $W_{i,t} = V_{i,t} = 0$ can be mean square stabilized over the memoryless white Gaussian multiple access channel if the systems' parameters $\{\lambda_1, \lambda_2\}$ satisfy the following inequalities*

$$\log(\lambda_1) < \frac{1}{2}\log\left(1 + \frac{P_1\left(1-\rho^{\star 2}\right)}{N}\right),$$
$$\log(\lambda_2) < \frac{1}{2}\log\left(1 + \frac{P_2\left(1-\rho^{\star 2}\right)}{N}\right), \tag{71}$$

where $\rho^\star$ is the root in the open interval $(0,1)$ of the following fourth order polynomial

$$\left(P_1\left(1-\rho^2\right)+N\right)\left(P_2\left(1-\rho^2\right)+N\right) = \left(P_1 + P_2 + 2\rho\sqrt{P_1 P_2} + N\right)N. \tag{72}$$

Proof. The proof is given in Sec. 4.3. ■

Remark 9. *It can be shown that for fully correlated initial states, i.e., $\rho_0 = 1$, the stability conditions are given by*

$$\log(\lambda_i) < \frac{1}{2}\log\left(1 + \frac{P_1 + P_2 + 2\sqrt{P_1 P_2}}{N}\right), \quad i = 1, 2.$$

Remark 10. *The terms on the right hand side in (71) correspond to the sum-rate optimal achievable rate pair for the two sources over the white Gaussian multiple-access channel with noiseless feedback [22]. The stability region in (71) is smaller than the capacity region in [22]. This is because to ensure second moment stability the coding scheme has to have at least double exponential error decay.*

Stability results for the broadcast channel In the broadcast channel there is a joint encoder with an output power constraint contrary to the multiple-access channel where the two encoders have individual power constraints. Therefore the joint encoder in the broadcast channel has freedom to tradeoff between the powers allocated to the transmission of the observed states of the two plants.

Theorem 13. *The two scalar LTI systems in (70) with $W_{i,t} = V_{i,t} = 0$ can be mean square stabilized over the memoryless white Gaussian broadcast channel if the systems' parameters $\{\lambda_1, \lambda_2\}$ satisfy the following inequalities*

$$\begin{aligned} \log(\lambda_1) &< \frac{1}{2}\log\left(\frac{D^\star(N+N_1+P)}{D^\star(N+N_1)+g^2P(1-\rho^\star)}\right), \\ \log(\lambda_2) &< \frac{1}{2}\log\left(\frac{D^\star(N+N_2+P)}{D^\star(N+N_2)+P(1-\rho^\star)}\right), \end{aligned} \tag{73}$$

where $D^\star = 1+g^2+2g\rho^\star$, $g \geq 0$, and $\rho^\star$ is the largest root in the open interval $(0,1)$ of the following polynomial

$$\begin{aligned} \rho = &-\Big(D(N\Sigma+N_1N_2)\rho - gP\Sigma(1-\rho^2)\Big)\times \\ &\Big(\Pi\left(D(N+N_1)+g^2P(1-\rho^2)\right)\left(D(N+N_2)+g^2P(1-\rho^2)\right)\Big)^{-\frac{1}{2}} \end{aligned} \tag{74}$$

where $\Pi = (P+N+N_1)(P+N+N_2)$ and $\Sigma = P+N+N_1+N_2$.

Proof. The proof is given in Sec. 4.3. ∎

Remark 11. *The terms on the right hand side in (73) is an achievable rate pair for the two decoders over the white Gaussian broadcast channel with noiseless feedback [23].*

Remark 12. *If the noise components $Z_{1,t}$ and $Z_{2,t}$ are zero in the broadcast channel model, then the two controllers receive the same signal and this setup is equivalent to having a joint controller. Therefore the stability region for the joint-sensor joint-controller case can be obtained by setting $N_1 = N_2 = 0$ in (73).*

The parameter g in (73) can tradeoff between the stabilizability of the two plants and thus we can obtain a stability region for the given channel parameters by increasing g from zero to less than infinity. Fig. 11 shows some examples of stability regions for $P = 10$. The solid line shows the boundary of the stability region when $N = 0$ and $N_1 = N_2 = 1$, the dashed line shows the boundary

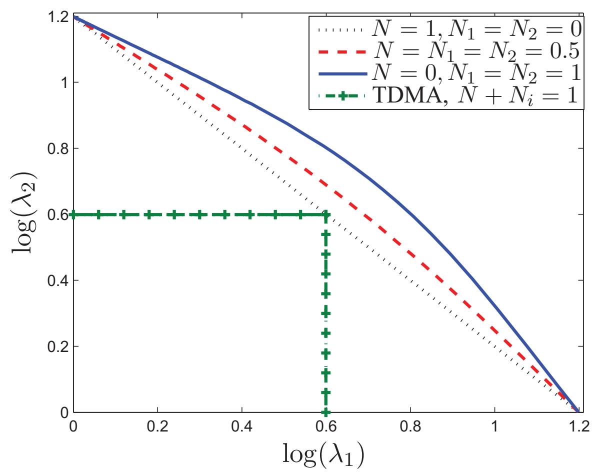


Figure 11: Illustration of the stability regions for the broadcast channel.

of the stability region when $N = N_1 = N_2 = 0.5$, and the dotted line shows the boundary of the stability region when $N = 1$ and $N_1 = N_2 = 0$. In these examples $N + N_i = 1$, and we can observe that the individual noise components $\{Z_{1,t}, Z_{2,t}\}$ are less harmful than the common noise component Z_t due to diversity effect. For comparison we also show the stability region when the encoder separately serves the two plants in alternate time steps, i.e., in each time step there is a point-to-point communication link from the encoder to one of the controllers. For this case the necessary and sufficient conditions for mean square stability can be found in [1], which are given by

$$\log(\lambda_i) < \frac{1}{4} \log\left(1 + \frac{P}{N + N_i}\right) \quad \text{for all } i \in \{1, 2\}.$$

The boundary of the rectangular stability region defined by the above inequalities is shown in Fig. 11 for $P = 10$ and $N + N_i = 1$.

Stability results for the Interference channel

Theorem 14. *The two scalar LTI systems in (70) with $W_{i,t} = 0$ can be mean square stabilized over the memoryless white Gaussian interference channel if the systems' parameters $\{\lambda_1, \lambda_2\}$ satisfy the following inequalities*

$$\log(\lambda_i) < \frac{1}{2} \log\left(\frac{P\left(1 + h^2 + 2h\rho^\star\right) + N}{Ph^2\left(1 - \rho^{\star 2}\right) + N}\right), \tag{75}$$

where $\rho^\star$ is the largest among all the roots in the interval $[0,1]$ of the following two fourth order polynomials

$$f_1(\rho) := \rho^4 + a_3\rho^2 + a_2\rho^2 + a_1\rho + a_0, \qquad (76)$$
$$f_2(\rho) := \rho^4 + b_3\rho^2 + b_2\rho^2 + b_1\rho + b_0,$$

where

$$a_3 = \frac{N}{2hP}, \quad a_2 = -2 - \frac{N(4+h\rho_z)}{2h^2P},$$
$$a_1 = -\frac{N(1+2h^2+2h\rho_z)}{2h^3P} - \frac{N^2}{h^3P^2},$$
$$a_0 = 1 + \frac{N(2h-\rho_z)}{2h^3P}, \quad b_3 = \frac{2h^2P+2P+N}{2hP},$$
$$b_2 = \frac{N\rho_z}{2hP}, \quad b_1 = -\frac{(1+h^2)}{h} - \frac{N(1+2\rho_z-2h^2)}{2h^3P},$$
$$b_0 = -1 - \frac{N(2h-\rho_z)}{2h^3P}.$$

Proof. The proof is given in Sec. 4.3. ■

Remark 13. *For fully correlated initial states, i.e., $\rho_0 = 1$, and fully correlated or anti-correlated noise components i.e., $\rho_z = \pm 1$, the initial transmissions in the proposed scheme in Sec. 4.3 can be modified such that $\rho^\star = 1$. Accordingly the stability conditions are then given by*

$$\log(\lambda_i) < \frac{1}{2}\log\left(1 + \frac{P(1+h)^2}{N}\right), \quad i = 1,2.$$

Remark 14. *It is shown in Appendix* **??** *that if the two noise components are fully correlated i.e., $\rho_z = 1$, and further $2h(1+\frac{h^2P}{N}) < 1$, then the largest root $\rho^\star$ of the polynomial $f_2(\rho)$ is equal to one. Therefore the stability conditions are then given by*

$$\log(\lambda_i) < \frac{1}{2}\log\left(1 + \frac{P(1+h)^2}{N}\right), \quad i = 1,2.$$

Remark 15. *The term on the right hand side in (75) corresponds to an achievable rate pair for the two sources over the white Gaussian interference channel with noiseless feedback [[5]].*

According to Theorem 14, the stabilizability of the two first order LTI systems depends on the given interference channel parameters such as average transmit power P, noise power N, noise cross-correlation ρ_z, and cross channel gain h. Therefore it is interesting to study the effect of these channel parameters on the behavior of the two systems under our proposed communication and control scheme. In this section we investigate the stabilizability of the two systems for different values of noise cross-correlation and cross channel gain with fixed transmit and noise powers.

In Fig. 12 we fix $P = 20$, $N = 1$, and plot the boundary of the stabilizability region for the two plants as a function of ρ_z for different values of h, according to Theorem 14. Therefore the ith plant will be mean square stable under our proposed scheme for the given channel parameters if $\log(\lambda_i)$ is in the region below the corresponding stabilizability boundary curve. In Fig. 12 we have shown some examples for different levels of interference parameter, i.e., $h = \{0, 0.15, 1, 100\}$. In most cases stabilizability region reduces by increasing ρ_z from -1 to 1, except for the case when the interference is very weak, i.e. $h = 0.15$. For this case, it is given in Remark 14 that the best performance is achieved when $\rho_z = 1$. For the sake of comparison we have also plotted no interference case ($h = 0$), i.e., there exist two parallel channels from the two plants to the two remote controllers. For this setup the stability conditions are given by $\log(\lambda_i) < 0.5 \log(1 + P/N)$. In Fig. 12 we have also shown an example of very strong interference scenario ($h = 100$). This example suggests that the stabilizability region significantly expands in the presence of a very strong interference. In order to investigate this further, we now fix $P = 20$, $N = 1$, and plot the boundary of the stabilizability region as a function of h for different values of ρ_z in Fig. 13. For $\rho_z = \{-1, -0.95\}$ the stabilizability increases monotonically with increasing cross channel gain h. Interestingly we observe a boost in the stabilizability of the systems for $\rho_z = -1$ compared to $\rho_z = -0.95$ in the high interference regime. For $\rho_z = \{0, 1\}$ the worst performance happens when the interference is moderate (i.e., neither weak nor strong). However in these examples too the stabilizability improves monotonically with increase in cross channel gain beyond certain threshold. Further we observe that in the low interference regime (i.e. for very small values of h) the best performance is achieved when the noise components are fully correlated $(\rho_z = 1)$, which is in accordance with Remark 14.

In the given setup of control over interference channel, the two systems are driven by the actions of the two controllers. These control actions are influenced by the cross channel interference, therefore it is also interesting to study cross-correlation between the state processes of the two systems for different values

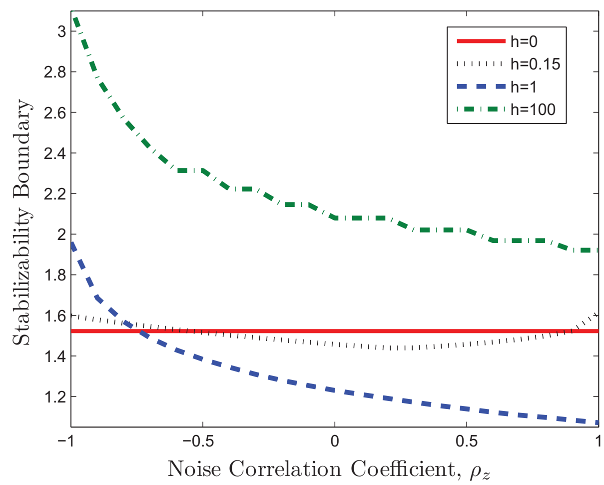


Figure 12: Examples of the stability region of the two systems as a function of cross-correlation coefficient of the two white noise components in the symmetric Gaussian interference channel, with fixed $P = 20$, $N = 1$.

of cross channel gain. Under our proposed scheme, the magnitude of the cross-correlation coefficient between the two state processes remains constant in the steady state, however it might alternate in phase in successive time steps as shown in Sec. 4.3. In Fig. 14 we fix $P = 20$, $N = 1$, and plot magnitude of the state cross-correlation coefficient $\rho^\star$ as a function of cross channel gain h for different values of noise correlation $\rho_z = \{-1, -0.5, 0, 1\}$. In these examples, we observe a general trend that cross correlation increases in magnitude as cross channel gain increases. The state processes of the two systems become almost fully correlated as cross channel interference gets very strong. This happens because the two systems are driven control actions which are highly influenced by cross talk.

In summary, the above numerical examples suggest that the stabilizability of the two systems over Gaussian interference channel reduces with the increase of noise cross-correlation from -1 to $+1$. That is negative correlation helps and positive correlation hurts except for the case when interference is very weak (see also Remark 14). Further for anti-correlated ($\rho_z = -1$) noises there is a dramatic boost in the stabilizability especially in the presence of strong interference. A similar behavior was observed in [[5]], where the authors showed that the sum-rate capacity over symmetric Gaussian interference channel can be doubled with feedback in high SNR when the noise components are anti-correlated. Furthermore we have observed that in general stabilizability improves as the interference

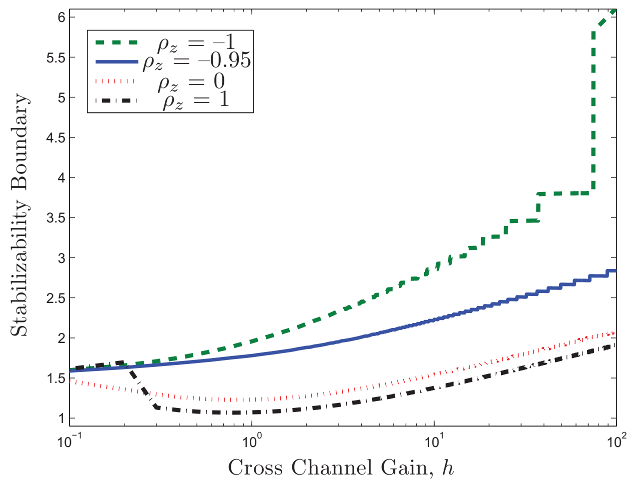


Figure 13: Examples of the stabilizability region of the two systems as a function of cross channel gain in the symmetric Gaussian interference channel, with fixed $P = 20$, $N = 1$.

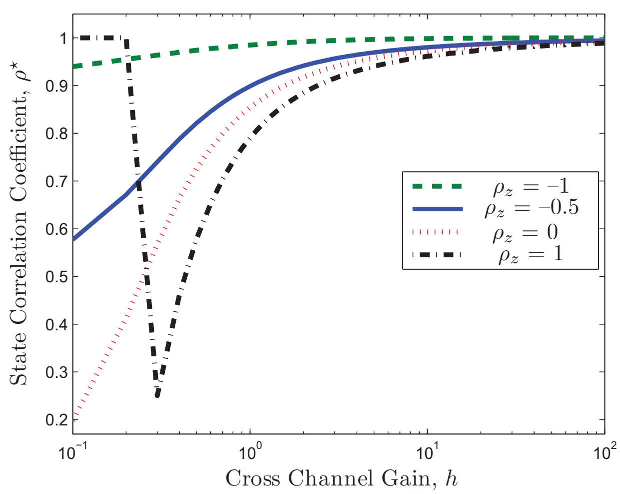


Figure 14: Illustration of cross correlation coefficient of the state processes of the two systems in steady state as a function of cross channel gain for different values of noise correlation, with fixed $P = 20$, $N = 1$.

gets significantly strong. This result is in line with already known results in information theory, where it has been shown that the transmission rates over interference channel can be significantly improved in presence of very strong interference [[4]]. Further we have observed that the state processes of the two systems become highly correlated in magnitude in strong interference scenarios under our proposed scheme.

The stability results provided in this paper can be extended for non-symmetric interference channel using the proposed scheme with some tedious computations. We can also extend our results for the setup where the links from the controllers to the plants are also white Gaussian communication channels. For this setup we can have an encoder at each control unit to encode the control action and a MMSE decoder at each plant to decode the transmitted value of the control action. As long as the encoders, the decoders, and the controllers are linear, the nature of the problem does not change and the stability results can be easily obtained cf. [[16]].

4.3 Control and Communication Schemes

In order to prove Theorems 12, 13, and 14, we propose to use coding schemes in [2, 5, 22, 23]. These schemes are based on Schalkwijk–Kailath coding scheme [118]. By employing the proposed coding schemes over the given broadcast and multiple-access channels, we then find conditions on the system parameters $\{\lambda_1, \lambda_2\}$ which are sufficient to mean square stabilize the systems in (70).

Scheme for the Multiple-access Channel The scheme for the white Gaussian multiple-access channel works as follows.

Initial time steps, $t = 0, 1$

Initially the two encoders transmit the observed state values in alternate time slots to the respective controllers. The first two disjoint transmissions in time make the plant states Gaussian distributed regardless of the distribution of their initial states, which will be explained shortly. However if the initial states are already Gaussian, then the following disjoint initial transmissions are not needed.

At time step $t = 0$, the encoder $\mathcal{E}_1$ observes $X_{1,0}$ and transmits $S_{1,0} = \sqrt{\frac{P_1}{\alpha_{1,0}}} X_{1,0}$. The encoder $\mathcal{E}_2$ stays quiet, i.e., $S_{2,0} = 0$. The decoder $\mathcal{D}$ receives $R_0 = S_{1,0} + Z_0$.

It then estimates $X_{1,0}$ as

$$\hat{X}_{1,0} = \sqrt{\frac{\alpha_{1,0}}{P_1}} R_0 = X_{1,0} + \sqrt{\frac{\alpha_{1,0}}{P_1}} Z_0.$$

The controller $\mathcal{C}$ then takes an action $U_{1,0} = -\lambda_1 \hat{X}_{1,0}$ for the plant 1, which results in $X_{1,1} = \lambda_1(X_{1,0} - \hat{X}_{1,0})$. The state $X_{1,1} \sim \mathcal{N}(0, \alpha_{1,1})$ with $\alpha_{1,1} = \lambda_1^2 \frac{\alpha_{1,0} N}{P_1}$. The controller does not take any action for the plant 2, therefore $X_{2,1} = \lambda_2 X_{2,0}$ with $\alpha_{2,1} = \lambda_2^2 \alpha_{2,0}$.

At time step $t = 1$, the encoder $\mathcal{E}_1$ stays quiet. The encoder $\mathcal{E}_2$ observes $X_{2,1}$ and transmits $S_{2,1} = \sqrt{\frac{P_2}{\alpha_{2,1}}} X_{2,1}$. The decoder $\mathcal{D}$ receives $R_1 = S_{2,1} + Z_1$. It then estimates $X_{2,1}$ as

$$\hat{X}_{2,1} = \sqrt{\frac{\alpha_{2,1}}{P_2}} R_1 = X_{2,1} + \sqrt{\frac{\alpha_{2,1}}{P_2}} Z_1.$$

The controller $\mathcal{C}$ then takes an action $U_{2,1} = -\lambda_2 \hat{X}_{2,1}$ for the plant 2, which results in $X_{2,2} = \lambda_2(X_{2,1} - \hat{X}_{2,1})$. The state $X_{2,2} \sim \mathcal{N}(0, \alpha_{2,2})$. For the plant 1, the controller does not take any action $U_{1,1} = 0$, therefore $X_{1,2} = \lambda_1 X_{1,1}$ and $X_{1,2} \sim \mathcal{N}(0, \alpha_{1,2})$.

It is noteworthy that due to non-overlapping initial transmissions by the two encoders, the states $X_{1,2}$ and $X_{2,2}$ are now zero mean Gaussian variables with correlation coefficient $\rho_2 = \frac{\mathbb{E}[X_{1,2}X_{2,2}]}{\sqrt{\alpha_{1,2}\alpha_{2,2}}}$ equal to zero[7]. Henceforth the two encoders will transmit their signals simultaneously.

Further time steps $t \geq 2$

The two encoders $\mathcal{E}_1$ and $\mathcal{E}_2$ observe $X_{1,t}$ and $X_{2,t}$, and they respectively transmit

$$S_{1,t} = \sqrt{\frac{P_1}{\alpha_{1,t}}} X_{1,t}, \quad S_{2,t} = \sqrt{\frac{P_2}{\alpha_{2,t}}} X_{2,t} \mathrm{sgn}(\rho_t),$$

where $\rho_t = \frac{\mathbb{E}[(X_{1,t} - \mathbb{E}[X_{1,t}])(X_{2,t} - \mathbb{E}[X_{2,t}])]}{\sqrt{\alpha_{1,t}\alpha_{2,t}}}$ and $\mathrm{sgn}(\rho_t) = 1$ if $\rho_t \geq 0$ and $\mathrm{sgn}(\rho_t) = -1$ if $\rho_t < 0$.

[7]The states in the second time step become uncorrelated irrespective of the value of the correlation between the initial states. This scheme does not exploit correlation between the initial states and thus the stability region obtained is independent of the correlation of the initial states.

The decoder $\mathcal{D}$ receives $R_t = S_{1,t} + S_{2,t} + Z_t$. It then computes an MMSE estimate of the state of the plant i as

$$\hat{X}_{i,t} = \mathbb{E}[X_{i,t}|R_t] \stackrel{(a)}{=} \frac{\mathbb{E}[R_t X_{i,t}]}{\mathbb{E}[R_t^2]} R_t, \tag{77}$$

where (a) follows from the fact that the optimum MMSE of the Gaussian variable is linear [34]; and we have

$$\begin{aligned}
\mathbb{E}\left[X_{1,t} R_t\right] &= \sqrt{\alpha_{1,t}} \left(\sqrt{P_1} + \sqrt{P_2}|\rho_t|\right), \\
\mathbb{E}[X_{2,t} R_t] &= \sqrt{\alpha_{2,t}} \left(\sqrt{P_2} + \sqrt{P_1}|\rho_t|\right) \operatorname{sgn}(\rho_t), \\
\mathbb{E}[R_t^2] &= P_1 + P_2 + 2|\rho_t|\sqrt{P_1 P_2} + N.
\end{aligned} \tag{78}$$

The controller $\mathcal{C}$ takes an action $U_{i,t} = -\lambda_i \hat{X}_{i,t}$ for the plant i, which results in $X_{i,t+1} = \lambda_i (X_{i,t} - \hat{X}_{i,t})$. The mean values of the states are

$$\begin{aligned}
\mathbb{E}[X_{i,t}] &= \mathbb{E}\left[\lambda_i \left(X_{i,t} - \hat{X}_{i,t}\right)\right] \\
&\stackrel{(a)}{=} \lambda_i \mathbb{E}\left[X_{i,t} - \frac{\mathbb{E}[R_t X_{i,t}]}{\mathbb{E}[R_t^2]} R_t\right] \stackrel{(b)}{=} 0,
\end{aligned} \tag{79}$$

where (a) follows from (77); and (b) follows from $\mathbb{E}[X_{i,2}] = 0$ and by recursively using (a). The variance of the state $X_{i,t+1}$ is given by

$$\begin{aligned}
\alpha_{i,t+1} \triangleq \mathbb{E}[X_{i,t+1}^2] &= \lambda_i^2 \mathbb{E}\left[\left(X_{i,t} - \hat{X}_{i,t}\right)^2\right] \\
&= \lambda_i^2 \mathbb{E}\left[\left(X_{i,t} - \frac{\mathbb{E}[R_t X_{i,t}]}{\mathbb{E}[R_t^2]} R_t\right)^2\right] \\
&= \lambda_i^2 \left(\mathbb{E}[X_{i,t}^2] - \frac{(\mathbb{E}[R_t X_{i,t}])^2}{\mathbb{E}[R_t^2]}\right).
\end{aligned} \tag{80}$$

By using (78) in (80) we get the following recursive equations

$$\alpha_{1,t+1} = \alpha_{1,t} \lambda_1^2 \left(\frac{N + P_2(1 - \rho_t^2)}{P_1 + P_2 + 2|\rho_t|\sqrt{P_1 P_2} + N}\right) \tag{81}$$

$$\alpha_{2,t+1} = \alpha_{2,t} \lambda_2^2 \left(\frac{N + P_1(1 - \rho_t^2)}{P_1 + P_2 + 2|\rho_t|\sqrt{P_1 P_2} + N}\right) \tag{82}$$

The cross-correlation between the states is given by

$$\begin{aligned}\mathbb{E}[X_{1,t+1}X_{2,t+1}] &= \mathbb{E}\left[\lambda_1\left(X_{1,t}-\hat{X}_{1,t}\right)\lambda_2\left(X_{2,t}-\hat{X}_{2,t}\right)\right]\\ &\overset{(a)}{=} \lambda_1\lambda_2\left(\mathbb{E}[X_{1,t}X_{2,t}]-\frac{\mathbb{E}[X_{1,t}R_t]\mathbb{E}[X_{2,t}R_t]}{\mathbb{E}[R_t^2]}\right)\\ &\overset{(b)}{=} \lambda_1\lambda_2\sqrt{\alpha_{1,t}\alpha_{2,t}}\left(\frac{N\rho_t-\text{sgn}(\rho_t)\sqrt{P_1P_2}(1-\rho_t^2)}{P_1+P_2+2|\rho_t|\sqrt{P_1P_2}+N}\right),\end{aligned} \tag{83}$$

where (a) follows from $\mathbb{E}[\hat{X}_{1,t}X_{2,t}]=\mathbb{E}[\hat{X}_{2,t}X_{1,t}]=\mathbb{E}[\hat{X}_{1,t}\hat{X}_{2,t}]=\frac{\mathbb{E}[X_{1,t}R_t]\mathbb{E}[X_{1,t}R_t]}{\mathbb{E}[R_t^2]}$; and (b) follows from (78). The correlation coefficient is then given by

$$\begin{aligned}\rho_{t+1} &= \frac{\mathbb{E}[X_{1,t+1}X_{2,t+1}]}{\sqrt{\alpha_{1,t}\alpha_{2,t}}}\\ &\overset{(a)}{=} \lambda_1\lambda_2\sqrt{\frac{\alpha_{1,t}\alpha_{2,t}}{\alpha_{1,t+1}\alpha_{2,t+1}}}\left(\frac{N\rho_t-\text{sgn}(\rho_t)\sqrt{P_1P_2}(1-\rho_t^2)}{P_1+P_2+2|\rho_t|\sqrt{P_1P_2}+N}\right)\\ &\overset{(b)}{=} \frac{N\rho_t-\text{sgn}(\rho_t)\sqrt{P_1P_2}(1-\rho_t^2)}{\sqrt{(N+P_1(1-\rho_t^2))\,(N+P_2(1-\rho_t^2))}} \quad \forall t\geq 2,\end{aligned} \tag{84}$$

where (a) follows from (83); and (b) follows from (81) and (82). It has been shown in [22] that for (84) there exists a $\rho^\star$ such that if $\rho_t=\rho^\star$ then $\rho_{t+k}=(-1)^k\rho^\star$ for all $k\geq 0$, where $\rho^\star$ is the root in the open interval $(0,1)$ of the following fourth order polynomial.

$$\begin{aligned}\left(P_1\left(1-\rho^2\right)+N\right)\left(P_2\left(1-\rho^2\right)+N\right) &=\\ &\left(P_1+P_2+2\rho\sqrt{P_1P_2}+N\right)N.\end{aligned} \tag{85}$$

If we modify the control actions such that ρ_2 becomes equal to $\rho^\star$ instead of zero, then ρ_t will be equal to $(-1)^t\rho^\star$ for all $t\geq 2$. Suppose in the time step $t=1$ the controller takes the actions $U_{1,1}=m$ and $U_{2,1}=-\lambda_2\hat{X}_{2,1}+m$, where m is a Gaussian variable with zero mean and variance σ_m^2. By varying σ_m^2 the correlation coefficient ρ_2 can be made equal to any value between zero and one. Therefore by choosing σ_m^2 such that $\rho_2=\rho^\star$, we can rewrite (81) and (82) as

$$\begin{aligned}\alpha_{i,t} &= \alpha_{i,2}\left(\lambda_i^2\frac{N+P_i(1-\rho^\star)}{P_1+P_2+2|\rho^\star|\sqrt{P_1P_2}+N}\right)^{t-2}\\ &= \alpha_{i,2}\left(\lambda_i^2\frac{N}{N+P_i(1-\rho^\star)}\right)^{t-2},\end{aligned} \tag{86}$$

where the last equality follows from (85). We observe from (86) that $\alpha_{i,t} \to 0$ as $t \to \infty$ if

$$\left(\lambda_i^2 \frac{N}{N + P_i(1-\rho^\star)}\right) < 1$$
$$\Rightarrow \log(\lambda_i) < \frac{1}{2}\log\left(1 + \frac{P_i\left(1-\rho^{\star 2}\right)}{N}\right),$$

which completes the proof. $\square$

Scheme for the Broadcast Channel The communication and control scheme for the white Gaussian broadcast channel is in principle similar to that of the multiple-access channel where in the beginning the encoder separately transmit the states of the two plants in order to make them Gaussian. Thereafter the Gaussian distributed states are transmitted jointly. This scheme works as follows.

Initial time steps, $t = 0, 1$

In the first two time steps the encoder transmits state observations of each plant separately. These separate initial transmissions make plant states Gaussian distributed regardless of the distribution of their initial states. However if the initial states are already Gaussian, then the following separate initial transmissions are not needed.

At time step $t = 0$ the encoder ignores $X_{2,0}$ and transmits $X_{1,0}$ as $S_0 = \sqrt{\frac{P}{\alpha_{1,0}}}X_{1,0}$. The decoder $\mathcal{D}_1$ receives $R_{1,0} = S_0 + Z_0 + Z_{1,0}$. It then estimates $X_{1,0}$ as

$$\hat{X}_{1,0} = \sqrt{\frac{\alpha_{1,0}}{P}}R_{1,0} = X_{1,0} + \sqrt{\frac{\alpha_{1,0}}{P}}(Z_0 + Z_{1,0}).$$

The controller $\mathcal{C}_1$ then takes an action $U_{1,0} = -\lambda_1\hat{X}_{1,0}$ for the plant 1, which results in $X_{1,1} = \lambda_1(X_{1,0} - \hat{X}_{1,0})$. The state $X_{1,1} \sim \mathcal{N}(0, \alpha_{1,1})$ with $\alpha_{1,1} = \lambda_1^2\frac{\alpha_{1,0}(N+N_1)}{P}$. The controller $\mathcal{C}_2$ does not take any action for the plant 2, therefore $X_{2,1} = \lambda_2 X_{2,0}$ with $\alpha_{2,1} = \lambda_2^2\alpha_{2,0}$.

At time step $t = 1$ the encoder $\mathcal{E}$ ignores $X_{1,1}$ and transmits only $X_{2,1}$, i.e., $S_1 = \sqrt{\frac{P}{\alpha_{2,1}}}X_{2,1}$. The decoder $\mathcal{D}_2$ receives $R_{2,1} = S_1 + Z_1 + Z_{2,1}$. It then estimates $X_{2,1}$ as

$$\hat{X}_{2,1} = \sqrt{\frac{\alpha_{2,1}}{P}}R_{2,1} = X_{2,1} + \sqrt{\frac{\alpha_{2,1}}{P}}(Z_1 + Z_{2,1}).$$

The controller $\mathcal{C}_2$ then takes an action $U_{2,1} = -\lambda_2 \hat{X}_{2,1}$ for the plant 2, which results in $X_{2,2} = \lambda_2(X_{2,1} - \hat{X}_{2,1})$. The state variable $X_{2,2} \sim \mathcal{N}(0, \alpha_{2,2})$. The controller $\mathcal{C}_1$ does not take any action for the plant 1, i.e., $U_{1,1} = 0$, therefore $X_{1,2} = \lambda_1 X_{1,1}$ and $X_{1,2} \sim \mathcal{N}(0, \alpha_{1,2})$.

Similar to the multiple-access channel, the states $X_{1,2}$ and $X_{2,2}$ are now zero mean Gaussian variables with correlation coefficient $\rho_2 = \frac{\mathbb{E}[X_{1,2}X_{2,2}]}{\sqrt{\alpha_{1,2}\alpha_{2,2}}}$ equal to zero[8]. Henceforth the encoder will serve both plants simultaneously.

Further time steps, $t \geq 2$

The encoder $\mathcal{E}$ observes $X_{1,t}$ and $X_{2,t}$, and it transmits

$$S_t = \sqrt{\frac{P}{D_t}} \left(\frac{X_{1,t}}{\sqrt{\alpha_{1,t}}} + g \frac{X_{2,t}}{\sqrt{\alpha_{2,t}}} \operatorname{sgn}(\rho_t) \right), \tag{87}$$

where $D_t = 1 + g^2 + 2g|\rho_t|$, $g \geq 0$, $\rho_t = \frac{\mathbb{E}[(X_{1,t} - \mathbb{E}[X_{1,t}])(X_{2,t} - \mathbb{E}[X_{2,t}])]}{\sqrt{\alpha_{1,t}\alpha_{2,t}}}$, and $\operatorname{sgn}(\rho_t) = 1$ if $\rho_t \geq 0$ and $\operatorname{sgn}(\rho_t) = -1$ if $\rho_t < 0$.
The decoder $\mathcal{D}_i$ receives $R_{i,t} = S_t + Z_t + Z_{i,t}$. It then computes an MMSE estimate of the state of the plant i as

$$\begin{aligned} \hat{X}_{i,t} &= \mathbb{E}\left[X_{i,t} | \{R_{i,k}\}_{k=0}^{t}\right] \\ &\overset{(a)}{=} \mathbb{E}[X_{i,t}|R_{i,t}] \overset{(b)}{=} \frac{\mathbb{E}[R_{i,t}X_{i,t}]}{\mathbb{E}[R_{i,t}^2]} R_{i,t}, \end{aligned} \tag{88}$$

where (a) follows from $\mathbb{E}[X_{i,t}R_{i,k}] = 0$ for all $k < t$ and $X_{i,t}$ and $R_{i,t}$ are Gaussian variables; (b) follows from the fact that the optimum MMSE of the Gaussian variable is linear [34]; and we have

$$\begin{aligned} \mathbb{E}\left[X_{1,t}R_{1,t}\right] &= \sqrt{\frac{P\alpha_{1,t}}{D_t}} \left(1 + g|\rho_t|\right) \\ \mathbb{E}[X_{2,t}R_{2,t}] &= \sqrt{\frac{P\alpha_{2,t}}{D_t}} \left(\rho_t + g \operatorname{sgn}\left(\rho_t\right)\right) \\ \mathbb{E}[R_{i,t}^2] &= P + N + N_i. \end{aligned} \tag{89}$$

[8]The states in the second time step become uncorrelated irrespective of the value of the correlation between the initial states. This scheme does not exploit correlation between the initial states and thus the stability region obtained is independent of the correlation of the initial states.

The controller $\mathcal{C}_i$ takes an action $U_{i,t} = -\lambda_i \hat{X}_{i,t}$ for the plant i, which results in $X_{i,t+1} = \lambda_i(X_{i,t} - \hat{X}_{i,t})$. The mean values of the states are

$$\begin{aligned}
\mathbb{E}[X_{i,t+1}] &= \mathbb{E}\left[\lambda_i \left(X_{i,t} - \hat{X}_{i,t}\right)\right] \\
&\stackrel{(a)}{=} \lambda_i \mathbb{E}\left[X_{i,t} - \frac{\mathbb{E}[R_{i,t}X_{i,t}]}{\mathbb{E}[R_{i,t}^2]} R_{i,t}\right] \stackrel{(b)}{=} 0,
\end{aligned}$$

where (a) follows from (88); and (b) follows from $\mathbb{E}[X_{i,2}] = 0$ and by recursively using (a). The variance of the state $X_{i,t+1}$ is given by

$$\begin{aligned}
\alpha_{i,t+1} \triangleq \mathbb{E}[X_{i,t+1}^2] &= \lambda_i^2 \mathbb{E}\left[\left(X_{i,t} - \hat{X}_{i,t}\right)^2\right] \\
&= \lambda_i^2 \left(\mathbb{E}[X_{i,t}^2] - \frac{(\mathbb{E}[R_{i,t}X_{i,t})^2]}{\mathbb{E}[R_{i,t}^2]}\right).
\end{aligned} \tag{90}$$

By using (89) in (90) we get the following recursive equations.

$$\alpha_{1,t+1} = \alpha_{1,t}\lambda_1^2 \left(\frac{D_t(N+N_1) + g^2 P(1-\rho_t^2)}{D_t(P+N+N_1)}\right) \tag{91}$$

$$\alpha_{2,t+1} = \alpha_{2,t}\lambda_2^2 \left(\frac{D_t(N+N_2) + P(1-\rho_t^2)}{D_t(P+N+N_2)}\right). \tag{92}$$

The cross-correlation between the states is given by

$$\begin{aligned}
&\mathbb{E}[X_{1,t+1}X_{2,t+1}] = \mathbb{E}\left[\lambda_1\left(X_{1,t} - \hat{X}_{1,t}\right)\lambda_2\left(X_{2,t} - \hat{X}_{2,t}\right)\right] \\
&= \lambda_1\lambda_2\Big(\mathbb{E}[X_{1,t}X_{2,t}] - 2\mathbb{E}[\hat{X}_{1,t}X_{2,t}] + \mathbb{E}[\hat{X}_{1,t}\hat{X}_{2,t}]\Big) \\
&\stackrel{(a)}{=} \lambda_1\lambda_2\left(\frac{\mathbb{E}[X_{1,t}X_{2,t}]\Pi - \mathbb{E}[R_{1,t}X_{1,t}]\mathbb{E}[R_{2,t}X_{2,t}]\Sigma}{\Pi}\right) \\
&\stackrel{(b)}{=} \lambda_1\lambda_2\sqrt{\alpha_{1,t}\alpha_{2,t}}\Bigg(\rho_t \\
&\qquad - \frac{P}{D_t\Pi}(\rho_t + g|\rho_t|\rho_t + g\,\mathrm{sgn}(\rho_t) + g\rho_t)\Sigma\Bigg),
\end{aligned} \tag{93}$$

where (a) follows from $\mathbb{E}[\hat{X}_{1,t}X_{2,t}] = \frac{\mathbb{E}[R_{1,t}X_{1,t}]\mathbb{E}[R_{2,t}X_{2,t}]}{P+N+N_1}$, $\mathbb{E}[X_{1,t}\hat{X}_{2,t}] = \frac{\mathbb{E}[R_{1,t}X_{1,t}]\mathbb{E}[R_{2,t}X_{2,t}]}{P+N+N_2}$, $\mathbb{E}[\hat{X}_{1,t}\hat{X}_{2,t}] = \frac{\mathbb{E}[R_{1,t}X_{1,t}]\mathbb{E}[R_{2,t}X_{2,t}](P+N)}{(P+N+N_1)(P+N+N_2)}$, $\Pi \triangleq (P+N+N_1)(P+N+N_2)$, and

$\Sigma \triangleq (P + N + N_1 + N_2)$; and (b) follows from (89). Now we can write a recursive equation for the correlation coefficient ρ_t by using (91), (92) and (93), as

$$\rho_{t+1} = \frac{\mathbb{E}[X_{1,t+1}X_{2,t+1}]}{\sqrt{\alpha_{1,t+1}\alpha_{2,t+1}}} = (D_t(N\Sigma + N_1N_2)\rho_t - gP\Sigma(1-\rho_t^2)\mathrm{sgn}(\rho_t))$$
$$\times \left(\Pi\left(D_t(N+N_1) + g^2P(1-\rho_t^2)\right)\left(D_t(N+N_2) + g^2P(1-\rho_t^2)\right)\right)^{-\frac{1}{2}}$$

It has been shown in [23] that for the above recursive equation there exists a $\rho^\star$ such that if $\rho_t = \rho^\star$ then $\rho_{t+k} = (-1)^k\rho^\star$ for all $k \geq 0$, where $\rho^\star$ is the largest root in the open interval $(0, 1)$ of the polynomial given in (74). If we modify our encoding scheme such that ρ_2 becomes equal to $\rho^\star$ instead of zero, then ρ_t will be equal to $(-1)^t\rho^\star$ for all $t \geq 2$. Suppose in the initial transmissions (i.e., $t = 0, 1$) the encoder transmits $S_0 = \sqrt{\frac{P}{\alpha_{1,0}}}X_{1,0} + m$ and $S_1 = \sqrt{\frac{P}{\alpha_{2,1}}}X_{2,1} + m$, where m is a Gaussian variable with zero mean and variance σ_m^2. In this way ρ_2 can take on value between zero and one by varying σ_m^2. Thus by choosing σ_m^2 such that $\rho_2 = \rho^\star$, we can rewrite (81) and (82) as

$$\alpha_{1,t} = \alpha_{1,2}\left(\lambda_1^2\frac{D^\star(N+N_1) + g^2P(1-\rho^{\star 2})}{D^\star(P+N+N_1)}\right)^{t-2} \tag{94}$$

$$\alpha_{2,t} = \alpha_{2,2}\left(\lambda_2^2\frac{D^\star(N+N_2) + P(1-\rho^{\star 2})}{D^\star(P+N+N_2)}\right)^{t-2} \tag{95}$$

Although in the modified encoding scheme we have violated the average power constraint for the first two transmissions, its effect can be neglected for infinite time horizon. We observe from (94) that $\alpha_{1,t} \to 0$ as $t \to \infty$ if

$$\left(\lambda_1^2\frac{D^\star(N+N_1) + g^2P(1-\rho^{\star 2})}{D^\star(P+N+N_1)}\right) < 1$$
$$\Rightarrow \log(\lambda_1) < \frac{1}{2}\log\left(\frac{D^\star(P+N+N_1)}{D^\star(N+N_1) + g^2P(1-\rho^{\star 2})}\right).$$

Similarly it follows from (95) that $\alpha_{2,t} \to 0$ as $t \to \infty$ if

$$\log(\lambda_2) < \frac{1}{2}\log\left(\frac{D^\star(P+N+N_2)}{D^\star(N+N_2) + P(1-\rho^{\star 2})}\right) \qquad \square$$

Scheme for the Interference Channel The control and communication scheme for the interference channel works as follows.

Initial time steps, $t = 0, 1$

Initially the two encoders transmit the observed state values in alternate time slots to the respective controllers. The first two disjoint transmissions in time make the plant states Gaussian distributed regardless of the distribution of their initial states, which will be explained shortly. However if the initial states are already Gaussian, then the following disjoint initial transmissions are not needed.

At time step $t = 0$, the encoder $\mathcal{E}_1$ observes $X_{1,0}$ and transmits $S_{1,0} = \sqrt{\frac{P}{\alpha_{1,0}}}X_{1,0}$. The encoder $\mathcal{E}_2$ does not transmit, i.e., $S_{2,0} = 0$. The decoder $\mathcal{D}_1$ receives $R_{1,0} = S_{1,0} + Z_{1,0}$. It then estimates $X_{1,0}$ as

$$\hat{X}_{1,0} = \sqrt{\frac{\alpha_{1,0}}{P}}R_{1,0} = X_{1,0} + \sqrt{\frac{\alpha_{1,0}}{P}}Z_{1,0}.$$

The controller $\mathcal{C}_1$ then takes an action $U_{1,0} = -\lambda_1 \hat{X}_{1,0}$ for the plant 1, which results in $X_{1,1} = \lambda_1(X_{1,0} - \hat{X}_{1,0})$. The state $X_{1,1} \sim \mathcal{N}(0, \alpha_{1,1})$ with $\alpha_{1,1} = \lambda_1^2 \frac{\alpha_{1,0}N}{P}$. The controller does not take any action for the plant 2, therefore $X_{2,1} = \lambda_2 X_{2,0}$ with $\alpha_{2,1} = \lambda_2^2 \alpha_{2,0}$.

At time step $t = 1$, the encoder $\mathcal{E}_1$ does not transmit any signal. The encoder $\mathcal{E}_2$ observes $X_{2,1}$ and transmits $S_{2,1} = \sqrt{\frac{P}{\alpha_{2,1}}}X_{2,1}$. The decoder $\mathcal{D}_2$ receives $R_{2,1} = S_{2,1} + Z_{2,1}$. It then estimates $X_{2,1}$ as

$$\hat{X}_{2,1} = \sqrt{\frac{\alpha_{2,1}}{P}}R_{2,1} = X_{2,1} + \sqrt{\frac{\alpha_{2,1}}{P}}Z_{2,1}.$$

The controller $\mathcal{C}_2$ then takes an action $U_{2,1} = -\lambda_2 \hat{X}_{2,1}$ for the plant 2, which results in $X_{2,2} = \lambda_2(X_{2,1} - \hat{X}_{2,1})$. The state $X_{2,2} \sim \mathcal{N}(0, \alpha_{2,2})$. For the plant 1, the controller does not take any action $U_{1,1} = 0$, therefore $X_{1,2} = \lambda_1 X_{1,1}$ and $X_{1,2} \sim \mathcal{N}(0, \alpha_{1,2})$.

It is noteworthy that due to non-overlapping initial transmissions by the two encoders, the states $X_{1,2}$ and $X_{2,2}$ are now zero mean Gaussian variables with correlation coefficient $\rho_2 = \frac{\mathbb{E}[X_{1,2}X_{2,2}]}{\sqrt{\alpha_{1,2}\alpha_{2,2}}}$ equal to zero[9]. Henceforth the two encoders will transmit their signals simultaneously.

[9]The states in the second time step become uncorrelated irrespective of the value of the correlation between the initial states. This scheme does not exploit correlation between the initial states and thus the stability region obtained is independent of the correlation of the initial states.

Further time steps $t \geq 2$

The two encoders $\mathcal{E}_1$ and $\mathcal{E}_2$ observe $X_{1,t}$ and $X_{2,t}$, and they respectively transmit

$$S_{1,t} = \sqrt{\frac{P}{\alpha_{1,t}}} X_{1,t}, \quad S_{2,t} = \sqrt{\frac{P}{\alpha_{2,t}}} X_{2,t} \mathrm{sgn}(\rho_t),$$

where $\rho_t = \frac{\mathbb{E}[(X_{1,t}-\mathbb{E}[X_{1,t}])(X_{2,t}-\mathbb{E}[X_{2,t}])]}{\sqrt{\alpha_{1,t}\alpha_{2,t}}}$ and $\mathrm{sgn}(\rho_t) = 1$ if $\rho_t \geq 0$ and $\mathrm{sgn}(\rho_t) = -1$ if $\rho_t < 0$.
In accordance, the decoder $\mathcal{D}_1$ receives $R_{1,t} = S_{1,t} + hS_{2,t} + Z_{1,t}$ and the decoder $\mathcal{D}_2$ receives $R_{2,t} = S_{2,t} + hS_{1,t} + Z_{2,t}$. The decoder $\mathcal{D}_i$ then computes a memoryless [10] MMSE estimate of the state of the plant i as

$$\hat{X}_{i,t} = \mathbb{E}[X_{i,t}|R_{i,t}] \overset{(a)}{=} \frac{\mathbb{E}[R_{i,t}X_{i,t}]}{\mathbb{E}[R_{i,t}^2]} R_{i,t}, \tag{96}$$

where (a) follows from the fact that the optimum MMSE of the Gaussian variable is linear [34]; and we have

$$\begin{aligned} \mathbb{E}\left[X_{1,t}R_{1,t}\right] &= \sqrt{P\alpha_{1,t}}\left(1 + h|\rho_t|\right), \\ \mathbb{E}[X_{2,t}R_{2,t}] &= \sqrt{P\alpha_{2,t}}\left(1 + h|\rho_t|\right)\mathrm{sgn}(\rho_t), \\ \mathbb{E}[R_{i,t}^2] &= P(1 + h^2 + 2h|\rho_t|) + N. \end{aligned} \tag{97}$$

The controller $\mathcal{C}_i$ takes an action $U_{i,t} = -\lambda_i \hat{X}_{i,t}$ for the plant i, which results in $X_{i,t+1} = \lambda_i(X_{i,t} - \hat{X}_{i,t})$. The mean values of the states are

$$\begin{aligned} \mathbb{E}[X_{i,t+1}] &= \mathbb{E}\left[\lambda_i\left(X_{i,t} - \hat{X}_{i,t}\right)\right] \\ &\overset{(a)}{=} \lambda_i \mathbb{E}\left[X_{i,t} - \frac{\mathbb{E}[R_{i,t}X_{i,t}]}{\mathbb{E}[R_{i,t}^2]} R_{i,t}\right] \overset{(b)}{=} 0, \end{aligned} \tag{98}$$

[10] The memoryless estimator is not optimal since the channel outputs are correlated. Therefore we expect that an improvement might be possible if we use full memory in the estimator. However the analysis becomes complicated by considering full LMMSE estimation.

where (a) follows from (96); and (b) follows from $\mathbb{E}[X_{i,2}] = 0$ and by recursively using (a). The variance of the state $X_{i,t+1}$ is given by

$$\begin{aligned}\alpha_{i,t+1} &\triangleq \mathbb{E}[X_{i,t+1}^2] = \lambda_i^2 \mathbb{E}\left[\left(X_{i,t} - \hat{X}_{i,t}\right)^2\right] \\ &= \lambda_i^2 \mathbb{E}\left[\left(X_{i,t} - \frac{\mathbb{E}[R_{i,t}X_{i,t}]}{\mathbb{E}[R_{i,t}^2]} R_{i,t}\right)^2\right] \\ &= \lambda_i^2 \left(\mathbb{E}[X_{i,t}^2] - \frac{(\mathbb{E}[R_{i,t}X_{i,t}])^2}{\mathbb{E}[R_{i,t}^2]}\right). \end{aligned} \tag{99}$$

By using (97) in (99) we get the following recursive equations

$$\alpha_{i,t+1} = \alpha_{i,t}\lambda_i^2 \left(\frac{Ph^2\left(1 - |\rho_t|^2\right) + N}{P(1 + h^2 + 2h|\rho_t|) + N}\right), \tag{100}$$

for $i \in \{1, 2\}$.

The cross-correlation coefficient ρ_t between the two state processes for all $t \geq 3$ is given by

$$\begin{aligned}\rho_{t+1} &= \frac{\mathbb{E}[X_{1,t+1}X_{2,t+1}]}{\sqrt{\alpha_{1,t+1}\alpha_{2,t+1}}} = \frac{1}{\sqrt{\alpha_{1,t+1}\alpha_{2,t+1}}} \mathbb{E}\left[\lambda_1\left(X_{1,t} - \hat{X}_{1,t}\right)\lambda_2\left(X_{2,t} - \hat{X}_{2,t}\right)\right] \\ &\overset{(a)}{=} \frac{\lambda_1\lambda_2}{\sqrt{\alpha_{1,t+1}\alpha_{2,t+1}}} \times \left(\mathbb{E}[X_{1,t}X_{2,t}] - \frac{\mathbb{E}[X_{1,t}R_{2,t}]\mathbb{E}[X_{2,t}R_{2,t}]}{\mathbb{E}[R_{2,t}^2]}\right. \\ &\left. - \frac{\mathbb{E}[X_{2,t}R_{1,t}]\mathbb{E}[X_{1,t}R_{1,t}]}{\mathbb{E}[R_{1,t}^2]} + \frac{\mathbb{E}[X_{1,t}R_{1,t}]\mathbb{E}[X_{2,t}R_{2,t}]\mathbb{E}[R_{1,t}R_{2,t}]}{\mathbb{E}[R_{1,t}]\mathbb{E}[R_{2,t}]}\right) \\ &\overset{(b)}{=} \lambda_1\lambda_2\sqrt{\frac{\alpha_{1,t}\alpha_{2,t}}{\alpha_{1,t+1}\alpha_{2,t+1}}} \\ &\times \left(\rho_t - 2\frac{P\text{sgn}(\rho_t)\left(h + |\rho_t|\right)\left(1 + |\rho_t|\right)}{P(1 + h^2 + 2h|\rho_t|) + N} + \frac{P\text{sgn}(\rho_t)\left(1 + h|\rho_t|\right)^2\left(2hP + P|\rho_t|(1 + h^2) + N\rho_z\right)}{\left(P(1 + h^2 + 2h|\rho_t|) + N\right)^2}\right) \\ &\overset{(c)}{=} \text{sgn}(\rho_t)\left(\frac{P(1 + h^2 + 2h|\rho_t|) + N}{Ph^2\left(1 - |\rho_t|^2\right) + N}\right) \\ &\times \left(|\rho_t| - 2\frac{P\left(h + |\rho_t|\right)\left(1 + |\rho_t|\right)}{P(1 + h^2 + 2h|\rho_t|) + N} + \frac{P\left(1 + h|\rho_t|\right)^2\left(2hP + P|\rho_t|(1 + h^2) + N\rho_z\right)}{\left(P(1 + h^2 + 2h|\rho_t|) + N\right)^2}\right) \\ &\overset{(d)}{=} \text{sgn}(\rho_t).g(\rho_t), \qquad \forall t \geq 2. \end{aligned} \tag{101}$$

In the computation of (101), (a) follows from

$$\mathbb{E}[X_{1,t}\hat{X}_{2,t}] = \frac{\mathbb{E}[X_{1,t}R_{2,t}]\mathbb{E}[X_{2,t}R_{2,t}]}{\mathbb{E}[R_{2,t}^2]},$$
$$\mathbb{E}[X_{2,t}\hat{X}_{1,t}] = \frac{\mathbb{E}[X_{2,t}R_{1,t}]\mathbb{E}[X_{1,t}R_{1,t}]}{\mathbb{E}[R_{1,t}^2]},$$
$$\mathbb{E}[\hat{X}_{1,t}\hat{X}_{2,t}] = \frac{\mathbb{E}[X_{1,t}R_{1,t}]\mathbb{E}[X_{2,t}R_{2,t}]\mathbb{E}[R_{1,t}R_{2,t}]}{\mathbb{E}[R_{1,t}^2]\mathbb{E}[R_{2,t}^2]}, \tag{102}$$

(b) follows from

$$\mathbb{E}[X_{1,t}R_{2,t}] = \sqrt{P\alpha_{1,t}}(h + |\rho_t|),$$
$$\mathbb{E}[X_{2,t}R_{1,t}] = \sqrt{P\alpha_{2,t}}(h + |\rho_t|)\text{sgn}(\rho_t),$$
$$\mathbb{E}[R_{1,t}R_{2,t}] = 2hP + P|\rho_t|(1 + h^2) + N\rho_z,$$

(c) follows from (100); and (d) follows by defining $g(\rho_t)$.

Now we wish to find conditions on the parameters $\{\lambda_1, \lambda_2\}$ which ensure mean square stability of the two systems in (70) over the given white Gaussian interference channel. In order to find the values of the parameters $\{\lambda_1, \lambda_2\}$ for which the variance of the two state processes given by (100) can be made equal to zero as time goes to infinity, we make use of the following lemma.

Lemma 3. *For the recursive equation in* (101) *there exists at least one* $\rho^\star \in [0, 1]$ *such that if* $|\rho_t| = \rho^\star$ *then* $|\rho_{t+k}| = \rho^\star$ *for all* $k \geq 0$*, where* $\rho^\star$ *is a root of one of the two polynomials* $\{f_1(\rho), f_2(\rho)\}$ *given in* (76)*. Further if* $\rho^\star$ *is a root of* $f_1(\rho)$ *then* $\rho_{t+k} = (-1)^k\rho^\star$*, and if* $\rho^\star$ *is a root of* $f_2(\rho)$ *then* $\rho_{t+k} = \rho^\star$ *for all* $k \geq 0$*.*

Proof. The proof can be found in [8, Appendix A]. ■

If we modify our encoding scheme such that ρ_2 becomes equal to $\rho^\star$ instead of zero, then $|\rho_t|$ will be equal to $\rho^\star$ for all $t \geq 2$. This modification[11] in the encoding scheme can be done as follows. Suppose in the initial transmissions (i.e., $t = 0, 1$) the two encoders transmit $S_{1,0} = \sqrt{\frac{P}{\alpha_{1,0}}}X_{1,0} + m$ and $S_{2,1} = \sqrt{\frac{P}{\alpha_{2,1}}}X_{2,1} + m$, where m is a Gaussian variable with zero mean and variance σ_m^2. In this way ρ_2

[11] At this point we modify the encoding scheme in order to artificially guarantee convergence of ρ_t to a fixed point. Numerical experiments suggest that ρ_t always converges to a fixed point starting from an arbitrary ρ_2, and this fixed point is unique.

can take on any value between zero and one by varying σ_m^2. Thus by choosing σ_m^2 such that $\rho_2 = \rho^\star$, we can rewrite (100) as

$$\begin{aligned}\alpha_{i,t+1} &= \alpha_{i,t}\lambda_i^2 \left(\frac{Ph^2\left(1-\rho_t^{\star 2}\right)+N}{P(1+h^2+2h\rho_t^\star)+N}\right)\\ &= \alpha_{i,2}\left(\lambda_i^2 \frac{Ph^2\left(1-\rho_t^{\star 2}\right)+N}{P(1+h^2+2h\rho_t^\star)+N}\right)^{t-2}. \end{aligned} \tag{103}$$

Although in the modified encoding scheme we have violated the average power constraint for the first two transmissions, its effect can be neglected for infinite time horizon. We observe from (103) that $\alpha_{i,t} \to 0$ as $t \to \infty$ if

$$\begin{aligned}&\left(\lambda_i^2 \frac{Ph^2\left(1-\rho_t^{\star 2}\right)+N}{P(1+h^2+2h\rho_t^\star)+N}\right) < 1\\ &\Rightarrow \log(\lambda_i) < \frac{1}{2}\log\left(\frac{P\left(1+h^2+2h\rho^\star\right)+N}{Ph^2\left(1-\rho^{\star 2}\right)+N}\right), \end{aligned} \tag{104}$$

for i in {1,2}. The term on the right hand side in (104) is a monotonically increasing function of $\rho^\star$, therefore we choose $\rho^\star$ to be the largest among all roots in $[0,1]$ of the two polynomials $\{f_1(\rho), f_2(\rho)\}$. The condition on λ_i in (104) guarantees *mean square* stability of the ith the open loop system if it is unstable i.e., $\lambda_i > 1$. For $\lambda_i < 1$ (i.e., $\log(\lambda_i) < 0$), the open loop system is self stable and the variance of the state process will converge to zero without any control actions in closed-loop. Therefore the sufficient conditions for mean square stability are given by (75). This completes the proof of Theorem 14. □

4.4 Conclusion

We study the problem of mean square stabilizing two discrete time scalar LTI systems in closed-loop via control over white Gaussian multiple-access, broadcast, and interference communication channels. We propose to use simple linear communication and control schemes which whiten the state process and make it Gaussian, and therefore the optimal decoding of the transmitted state values at the remote control unit(s) is linear and memoryless. The stability regions obtained are associated with the achievable rate regions for the given channels with noiseless feedback. Therefore our results reveal relationship between mean square stability of the two plants and the communication channels' parameters, i.e., average power consumed by the encoder(s) and the average power of the noise components in different links.

The stability results provided in this paper can be easily extended for the setup where the links from the controller(s) to the plants are also white Gaussian communication channels. For this setup we can have an encoder at each control unit to encode the control action and an MMSE decoder at each plant to decode the transmitted value of the control action. As long as the encoders, the decoders, and the controllers are linear, the nature of the problem does not change and the stability results can be easily obtained [15].

5 Optimized encoder–controller mappings for closed-loop control over bandlimited noisy channels

In this section, we study iterative design of encoder–controller mappings for a closed-loop linear system with state feedback transmitted over a noisy channel. With the objective to minimize the expected linear quadratic cost over a finite horizon, we propose a joint design of the sensor measurement quantization, channel error protection, and controller actuation. It is argued that despite that this encoder–controller optimization problem is known to be hard in general, an iterative design procedure can be derived in which the controller is optimized for a fixed encoder, then the encoder is optimized for a fixed controller, etc.

Most work on control with limited information has been devoted to stability, while optimal designs are much less explored in the literature. Exceptions include the study of optimal stochastic control over communication channels, e.g., [83, 103, 124].

Our main concern is optimal average performance over a finite horizon, given a fixed data rate. In [54], we introduced an *iterative design procedure* for finding encoder–controller pairs. The result is a synthesis technique for joint optimization of the quantization, error protection and control over a bandlimited and noisy channel. This is an important problem in networked control in the case when a large set of sensor nodes need to limit their individual access to the communication medium.

Some notation used throughout this subsection is as follows. Bold-faced characters are used for describing a sequence of signals or functions, e.g., $\mathbf{x}_a^b = \{x_a, \ldots, x_b\}$ denotes the evolution of a discrete-time signal x_t from $t = a$ to $t = b$.

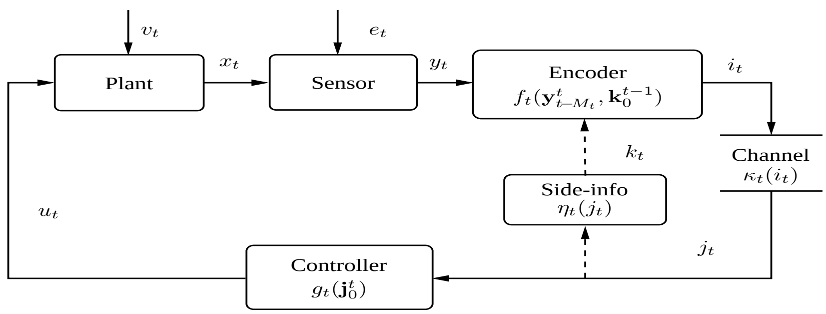


Figure 15: A general system for feedback control over a discrete memoryless channel. The dashed line indicates potential side-information signaling from the controller to the encoder.

We use $\mathbf{E}\{\cdot\}$ to denote the expectation operator. The notation $(\cdot)'$ stands for matrix transpose and $(\cdot)^\dagger$ matrix pseudoinverse. To indicate an optimal solution, the notation $(\cdot)^*$ is used.

5.1 Preliminaries

Consider the control system with a communication channel depicted in Fig. 15. The multi-variable linear plant is governed by the following equations

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t + v_t, \\ y_t &= Cx_t + e_t, \end{aligned} \tag{105}$$

where $x_t \in \mathbb{R}^n$, $u_t \in \mathbb{R}^m$, $y_t \in \mathbb{R}^p$ are the state, the control, and the measurement, respectively. The variable $v_t \in \mathbb{R}^n$ is the process noise and $e_t \in \mathbb{R}^p$ is the measurement noise. The noise signals are independent and identically distributed (i.i.d.) and mutually independent. They are also independent of the system initial state x_0.

The *encoder* is a mapping from the set of the encoder information to a discrete set of symbols. We take each symbol to be represented by an integer index. At time t, the index is $i_t \in \mathcal{CL}_i = \{1, \ldots, L_i\}$. In particular, we are interested in the class of encoder mappings described by the function

$$i_t = f_t(\mathbf{y}_{t-M_t}^t, \mathbf{k}_0^{t-1}), \tag{106}$$

where the parameter M_t specifies how many of the past measurements can be used by the encoder. Given the sequence of past side-information, $\mathbf{k}_0^{t-1}$, and the measurements, $\mathbf{y}_{t-M_t}^t$, the encoder produces an index i_t. The side-information k_t represents available feedback to the encoder about the value of the symbol $j_t \in \mathcal{C}L_j = \{1, \ldots, L_j\}$ received at the controller. We define the *side-information* (SI) at the encoder to be produced as

$$k_t = \eta_t(j_t) \in \mathcal{C}L_k = \{1, \ldots, L_k\}, \;\; 1 \leq L_k \leq L_j, \tag{107}$$

where $\eta_t : \mathcal{C}L_j \to \mathcal{C}L_k$ is deterministic and memoryless.

The encoder output indexes, i_t, are transmitted over a *discrete memoryless channel* (DMC), with input and output alphabets $\mathcal{C}L_i$ and $\mathcal{C}L_j$, respectively. The transmitted index is then received as j_t. One use of the channel is defined as

$$j_t = \kappa_t(i_t), \tag{108}$$

where $\kappa_t : \mathcal{C}L_i \to \mathcal{C}L_j$ is a random memoryless mapping. By assuming $L_j \geq L_i$, the output alphabet is potentially larger than the input alphabet, and hence we allow the possibility of soft information at the channel output.

At the receiver side, we consider a controller that causally utilizes all the available *controller information* $\mathbf{j}_0^t$ to produce the control command

$$u_t = g_t(\mathbf{j}_0^t) \in \mathbb{R}^m. \tag{109}$$

We denote $\hat{x}_{s|t} = \mathbf{E}\left\{x_s|\mathbf{j}_0^t\right\}$, and use $\hat{x}_t$ as a short notation for $\hat{x}_{t|t} = \mathbf{E}\left\{x_t|\mathbf{j}_0^t\right\}$. Then, $\tilde{x}_t = x_t - \hat{x}_t = x_t - \mathbf{E}\left\{x_t|\mathbf{j}_0^t\right\}$ is the estimation error.

Our goal is to solve an optimal encoder–controller problem and thereby finding the suitable encoder and controller mappings. The adopted performance measure is the LQ cost $\mathbf{E}\left\{J_T\right\}$, where

$$J_T = \sum_{t=1}^{T} \left(x_t' V_t x_t + u_{t-1}' P_{t-1} u_{t-1} \right). \tag{110}$$

The weighting matrices V_t and P_t are symmetric and positive definite.

Summarizing the above discussions, Problem 1 below specifies the encoder–controller optimization problem studied in this work.

Problem 1. *Consider the system in Fig. 15. Given the linear plant* (105) *and the memoryless channel* (108)*, find the encoder* (106) *and controller* (109) *that minimize the cost* $\mathbf{E}\left\{J_T\right\}$.

5.2 Iterative design

In general, finding an exact solution to Problem 1 is not feasible, because the optimization problem is highly non-linear and non-convex. Therefore, we propose a method to design the encoder–controller pair iteratively, with the goal of finding locally optimal solutions. Inspired by traditional quantizer and vector quantizer designs [78,86], the idea is to fix the encoder and update the controller, then fix the controller and update the encoder, etc. The iteration terminates when convergence is reached. Unfortunately, the iterative optimization algorithm will not guarantee convergence to a global optimum, but by influencing the initial conditions of the design it is possible to search for good locally optimal designs. Next, we describe the criteria used to update the controller mapping and encoder mapping, respectively.

Optimal Controller for Fixed Encoder

The problem of finding the optimal control assuming the encoder is fixed fits well into the setting of stochastic optimal control, e.g., [50]. We apply dynamic programming to derive the optimal control strategy recursively. Resembling a classical result in LQ control, we present the following proposition.

Proposition 3. *Consider a fixed encoder* $\mathbf{f}_0^{T-1}$. *Given the plant* (105) *and the memoryless channel* (108), *a controller mapping* (109) *that minimizes the LQ cost* $\mathbf{E}\{J_T\}$ *fulfills the following recursive relation*

$$\begin{aligned} u_{t-1}^* &= \arg\min_{u_{t-1}}\{\gamma_t\},\\ \gamma_t &= \lambda_t + \mathbf{E}\left\{\gamma_{t+1}^* | \mathbf{j}_0^{t-1}\right\},\\ \lambda_t &= \mathbf{E}\Big\{(Ax_{t-1} + Bu_{t-1} + v_{t-1})' V_t (Ax_{t-1} + Bu_{t-1} + v_{t-1}) + u_{t-1}' P_{t-1} u_{t-1} | \mathbf{j}_0^{t-1}\Big\}, \end{aligned} \tag{111}$$

for $t = 1, \ldots, T$, *where the cost-to-go* γ_t *is initialized with the optimal cost-to-go of* $t = T+1$, *namely,* $\gamma_{T+1}^* = 0$.

Unfortunately, it is in general not possible to efficiently solve (111). However, at the special case that the encoder has full SI, $k_t = j_t$, we are able to provide a characterization of the optimal system. As detailed in [54], the basic idea is that the closed-loop system can be converted into an equivalent open-loop encoder system, see Fig. 16. The detailed description of the open-loop encoder system can be found in [54]. Given the plant, the memoryless channel and the design criterion, the solutions to the original optimization problem and the corresponding

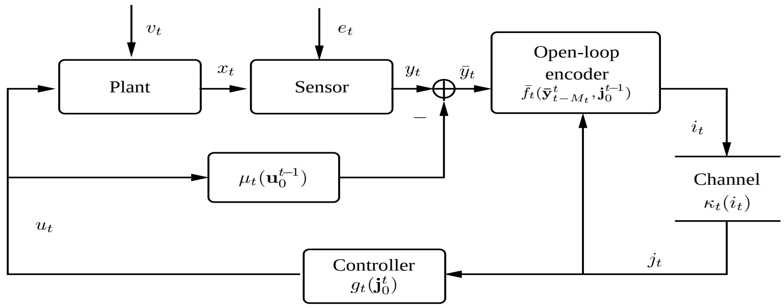


Figure 16: The open-loop encoder system, *a virtual help-system.*

optimization problem for the open-loop encoder system are closely related. Moreover, when using the iterative design approach, the open-loop encoder system is in general easier to deal with than the original system. Hence, in the special case of full SI we will focus on finding a solution to the open-loop encoder system, and then derive a corresponding solution to the original system. Consider an open-loop encoder system, we will be able to solve (111), as revealed by the following proposition.

Proposition 4. *Consider the open-loop encoder system, assuming a fixed open-loop encoder. Given the plant* (105) *and the memoryless channel* (108)*, the controller component* $u_t = g_t(\mathbf{j}_0^t)$ *that minimizes the LQ cost* $\mathbf{E}\{J_T\}$ *is given by*

$$u_t = \ell_t \hat{x}_t, \tag{112}$$

where $\hat{x}_t = \mathbf{E}\{x_t | \mathbf{j}_0^t\}$*. The control gain* ℓ_t *can be recursively computed as*

$$\begin{aligned} \ell_t &= -(P_t + B'(V_{t+1} + K_{T-t-1})B)^{\dagger} B'(V_{t+1} + K_{T-t-1})A, \\ K_{T-t-1} &= A'(V_{t+2} + K_{T-t-2})A - \pi_{T-t-1}, \\ \pi_{T-t-1} &= A'(V_{t+2} + K_{T-t-2})B(P_{t+1} + B'(V_{t+2} \\ &\qquad + K_{T-t-2})B)^{\dagger} B'(V_{t+2} + K_{T-t-2})A, \end{aligned} \tag{113}$$

where K_t *is initialized with* $K_1 = A'V_T A - A'V_T B(P_{T-1} + B'V_T B)^{\dagger} B'V_T A$*.*

The results in (112) and (113) illustrate that given a fixed open-loop encoder $\bar{\mathbf{f}}_0^{T-1}$, it is possible to explicitly characterize the optimal control strategy (111). Observe that the optimal control strategy (112) can be decomposed into a separate estimator/decoder and a controller. Hence, the *separation property* holds, e.g.,

[50]. Additionally, one can show that the derived optimal control strategy (112) is a *certainty equivalence* (CE) controller. In the optimal encoder–controller pair for the original system, the controller may not be separated without loss in the case of partial SI. Since we are not able to solve (111) in the general case, we resort to using the CE controller as a sub-optimal alternative to solving (111).

Optimal Encoder for Fixed Controller

The optimal encoder mapping needs to take the impact of the predicted future state evolutions into account. The following result is a straightforward consequence of the system assumptions and the design criterion.

Proposition 5. *Consider a fixed controller* $\mathbf{g}_0^{T-1}$ *and fixed encoder components* $\mathbf{f}_0^{t-1}$, $\mathbf{f}_{t+1}^{T-1}$. *Given the linear plant* (105) *and the memoryless channel* (108)*, the encoder component* $f_t(\mathbf{y}_{t-M_t}^t, \mathbf{k}_0^{t-1})$ *that minimizes the LQ cost* $\mathbf{E}\{J_T\}$ *is given by*

$$i_t = \arg\min_{i \in \mathcal{C}L_i} \mathbf{E}\Big\{ \sum_{s=t+1}^{T} (x_s' V_s x_s + u_{s-1}' P_{s-1} u_{s-1}) \Big| \mathbf{y}_{t-M_t}^t, \mathbf{k}_0^{t-1}, i_t = i \Big\}. \tag{114}$$

The encoder is specified by the encoder regions $\mathcal{C}S_i(\mathbf{k}_0^{t-1})$, $i \in \mathcal{C}L_i$, $t = 0, \ldots, T-1$. In the scalar case, $M_t = 0$ and $p = 1$, the regions can be specified by storing the boundaries between them.

Finally, we describe the encoder–controller design algorithm based on the above discussion. As mentioned, the overall joint encoder–controller optimization problem is typically not tractable, and we therefore propose to optimize the encoder–controller pair iteratively. There are two cases to handle separately:

1. *Full SI:* In this case, we carry out the design for the open-loop encoder system in Fig. 16 and then convert the solution to the original problem in Fig. 15.

2. *Partial SI:* In this case, we constrain the controller to be a CE controller, and carry out the design for the original system in Fig. 15.

Fig. 17 depicts a flow-diagram of the design procedure. First, an initial encoder–controller pair is specified. Thereafter, each component, f_0, g_0, $\ldots$, f_{T-1}, g_{T-1}, is successively optimized. After one round, if the improvement is not below a pre-defined threshold δ, a new round is started to update $f_0, g_0, \ldots, f_{T-1}, g_{T-1}$.

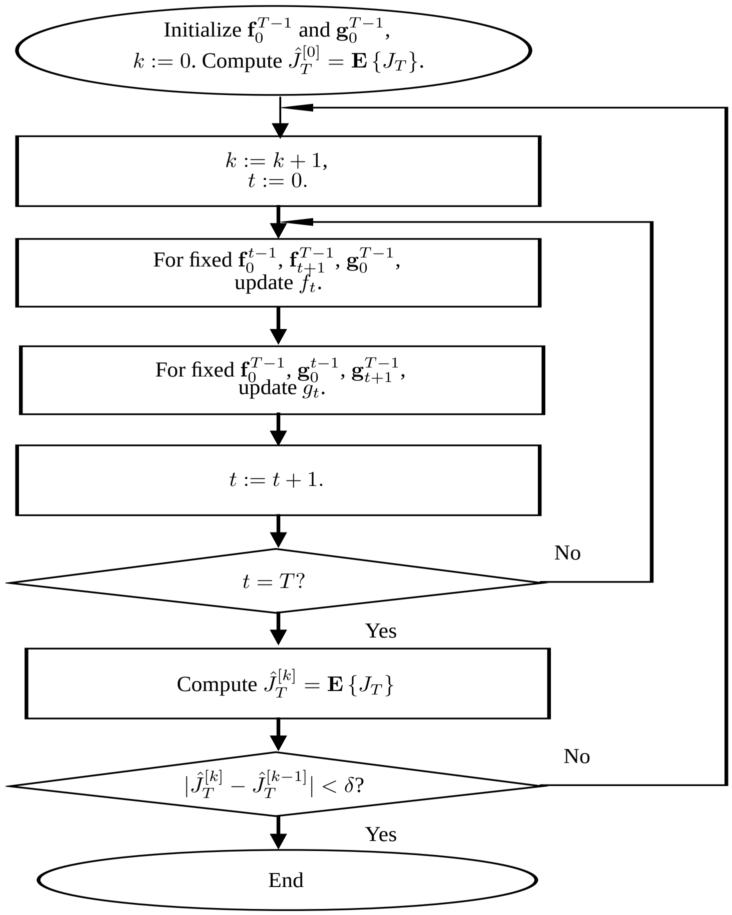


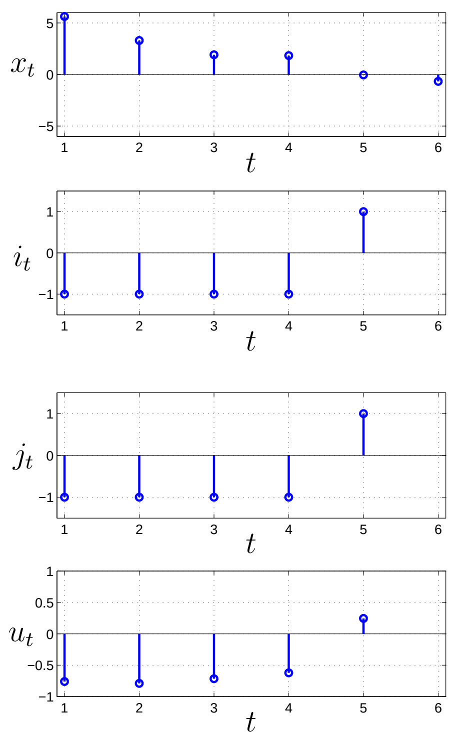
Figure 17: The flow-diagram of the iterative encoder–controller optimization procedure. The variable k is a counter for the number of rounds. In each round, all the encoder–controller components $\mathbf{f}_0^{T-1}$ and $\mathbf{g}_0^{T-1}$ are updated. The value $\hat{J}_T^{[k]}$ represents the resulting cost $\mathbf{E}\{J_T\}$ after round k. The iteration is terminated when the improvement in the system performance is less than a certain threshold δ.

5.3 Numerical examples

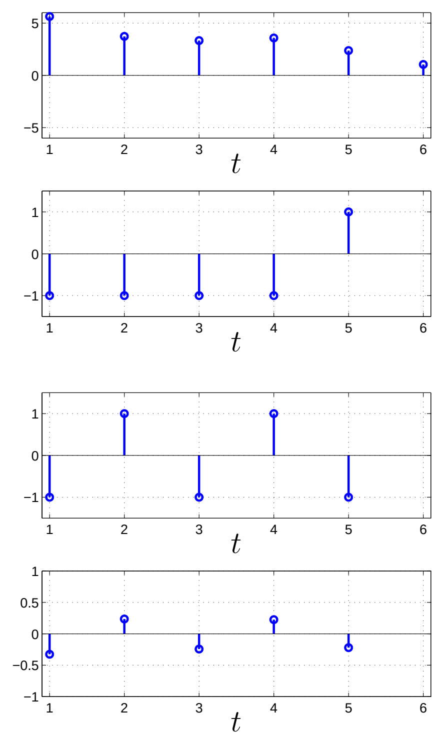
Here we present numerical experiments to demonstrate the performance obtained by using the iterative encoder–controller design. In particular, we study the case that encoded measurements are transmitted over a binary symmetric channel (BSC) for which the crossover probability is denoted by ϵ. In Fig. 18, the state response is depicted together with the transmitted symbol, the received symbol and the control. The system has been studied for the crossover probabilities $\epsilon = 0.04$ and 0.3. It can be observed that the number of symbol errors increases with ϵ. Since a symbol error can result in a control input doing more harm than good, as expected, the magnitude of the admissible control becomes smaller when the channel error increases.

In Fig. 19, we show the system performance in terms of the crossover probability ϵ. Performance $\bar{J}_T$ is obtained by normalizing $\mathbf{E}\{J_T\}$ with the expected cost obtained when no control action is taken, cf., the horizontal line in Fig. 19. Three types of encoder–controller pairs are studied: our proposed encoder–controller is compared with two heuristic designs A and B. The first pair, encoder–controller A, is designed as follows. The measurement is quantized using a time-invariant uniform quantizer. At the controller, received indexes are mapped into reconstructions which are fed into a Kalman filter for estimating the state x_t. The Kalman filter is designed assuming the error due to measurement noise, quantization and transmission is white and Gaussian distributed. Thereafter, the control is calculated as a linear function of the Kalman filter output. The linear feedback law is ℓ_t in (113). The second pair, encoder–controller B, utilizes a time-invariant uniform encoder, together with the CE controller in (112). The last pair is an encoder–controller trained according to our proposed design where the encoder has full SI. It can be seen in the figure that the trained encoder–controller pair outperforms the other two schemes.

How the encoder and controller respond to increasing the channel noise is illustrated in Fig. 20, using the same experiment setting as in Fig. 19. In the figure, we demonstrate the partition of the real numbers defined by the encoder mapping f_0, and the corresponding reconstructions $\hat{x}_0$, for growing ϵ. Recall that the control is a linear function of the reconstruction. We note that the number of controls chosen by the encoder decreases with increasing ϵ. This phenomenon is well-known in quantization for noisy channels and is attributed to the varying abilities of binary codewords in combating channel errors. For very noisy channels, it is beneficial to transmit only the "stronger" codewords, providing true redundancy for error protection. Note that the asymmetry at $\epsilon = 0.16$ is also a consequence



(a) $\epsilon = 0.04$



(b) $\epsilon = 0.3$

Figure 18: The system behavior is illustrated with respect to the crossover probability ϵ. The state response x_t, the transmitted symbol i_t, the received symbol j_t and the control u_t are depicted. In this example, $\epsilon = 0.04$ results in no transmission error and $\epsilon = 0.3$ in three errors.

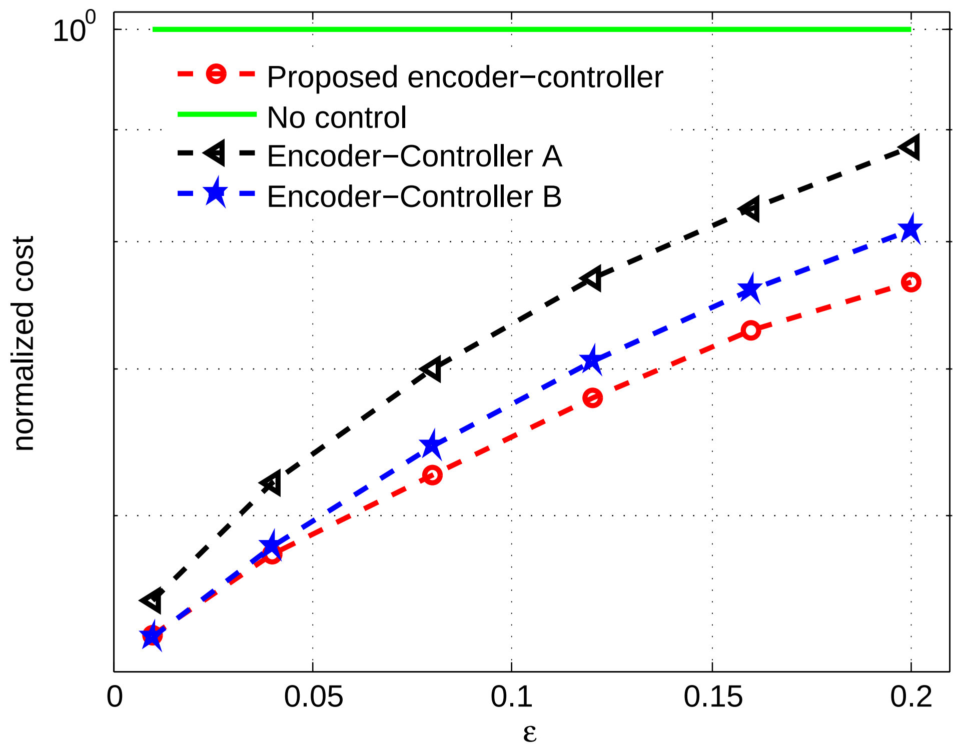


Figure 19: A comparison of the control performance between the proposed encoder–controller proposed and two heuristic encoder–controllers (encoder–controller A and encoder–controller B). Independent of the crossover probability, the proposed encoder–controller gives best performance.

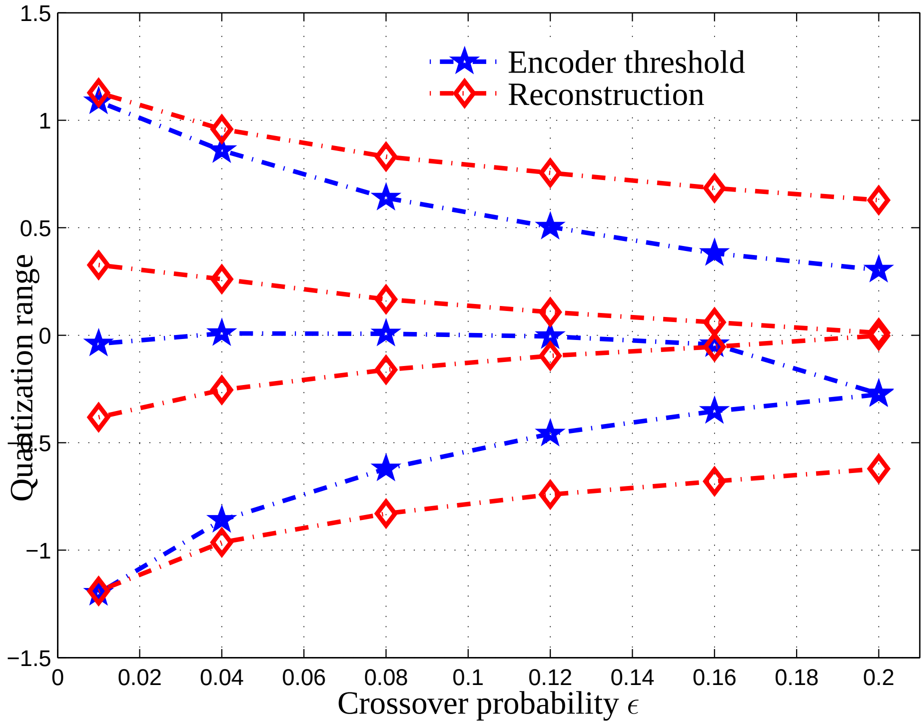


Figure 20: The encoding boundaries given by f_0 and the corresponding reconstructions $\hat{x}_0$ with respect to ϵ.

of the fact that the binary codewords are unequally sensitive to the channel errors. Another impact of increasing ϵ is that the encoder thresholds and the controls are all moved closer to zero, indicating that only small-valued control actions are allowed.

6 Consensus over noisy digital channels

In this section we will propose the coding techniques for consensus algorithms in the case of noisy digital channels introduced in [64], [65]. This methodology has already been described in details in the Deliverable D02.03 - Communication Network Design release 1 where its advantages in solving the consensus problem over noisy digital channels have been emphasized. Here we recall briefly the algorithm, and focus on its performance in terms of computational and communication complexities.

We consider a finite set of agents $\mathcal{V}$ of cardinality n and assume that each agent $v \in \mathcal{V}$ has access to some partial information consisting in the observation of a scalar value θ_v. The full vector of observations is denoted by $\boldsymbol{\theta} = (\theta_v)_{v\in\mathcal{V}}$. We consider the case when all θ_v's take values in the same bounded interval $\Theta \subseteq \mathbb{R}$. Such an interval may represent the common measurement range of the agents, possibly dictated by technological constraints, and assumed to be known a priori to all the agents. For ease of exposition, we assume that Θ coincides with the unitary interval $[0, 1]$. For the network, the goal is to compute the average of such values,

$$y := f(\boldsymbol{\theta}) = n^{-1} \sum_{v\in\mathcal{V}} \theta_v$$

through repeated exchanges of information among the agents and without a centralized computing system. Communication among the agents takes place as follows. At each time instant $t = 1, 2, \ldots$, every agent v broadcasts a binary signal $a_v(t) \in \{0, 1\}$ to its out-neighbourhood $\mathcal{N}_v^+$. Every agent $w \in \mathcal{N}_v^+$ receives a possibly erased version $b_{v\to w}(t) \in \{0, 1, ?\}$ of $a_v(t)$. Here, the symbol $?$ represents a lost binary signal. We denote by $b_v(t) = (b_{w\to v}(t))_{w\in\mathcal{N}_v^-}$, and $b'_v(t) = (b_{v\to w}(t))_{w\in\mathcal{N}_v^+}$ the vector of signals received by agent v at time t, and, respectively, the vector of signals received from agent v by its out-neighbours. At time t, each agent $v \in \mathcal{V}$ makes an estimate $\hat{y}_v(t)$ of y. The compact notation $\boldsymbol{a}(t) = (a_v(t))_{v\in\mathcal{V}}$, $\boldsymbol{b}(t) = (b_v(t))_{v\in\mathcal{V}}$, and $\hat{\boldsymbol{y}}(t) = (\hat{y}_v(t))_{v\in\mathcal{V}}$, is used for the full vectors of transmitted signals, received signals, and estimates at time t, respectively.

We assume the communication network to be memoryless, i.e., that $\boldsymbol{b}(t)$ is conditionally independent from the initial observations $\boldsymbol{\theta}$ and the previous transmissions $\{\boldsymbol{a}(s), \boldsymbol{b}(s) : 1 \leq s < t\}$, given the currently broadcasted signals $\boldsymbol{a}(t)$. Further, we assume that, given $\boldsymbol{a}(t)$, for every $v \in \mathcal{V}$ and $w \in \mathcal{N}_v^+$,

$$b_{v \to w}(t) = \begin{cases} ? & \text{w.p.} \quad \varepsilon \\ a_v(t) & \text{w.p.} \;\; 1 - \varepsilon. \end{cases}$$

Here ε is some erasure probability which, for simplicity, is assumed to remain constant in t, v and w. Distributedness of the computation algorithm is then modeled by constraining the transmitted signal $a_v(t)$ to be a function of the local information available to agent v at the end of the $(t-1)$-th round of communication, and the estimate $\hat{y}_v(t)$ to be a function of the information available to agent v at the end of the t-th round of communication. The local information available to agent v at the end of the t-th round of communication, consists of its initial observation, as well as of the signals received by v up to time t:

$$i_v(t) := \{\theta_v, b_v(s) : 1 \leq s \leq t\} \; .$$

Observe that the case $\varepsilon = 0$ reduces to one-bit-quantized transmission, which has been already considered in the literature.

The communication setting outlined above can be conveniently described by a directed graph $\mathcal{G}_\varepsilon = (\mathcal{V}, \mathcal{E})$, whose vertices are the agents, and such that an ordered pair (v, w) with $v \neq w$ belongs to $\mathcal{E}$ if and only if $w \in \mathcal{N}_v^+$ (or, equivalently, if $v \in \mathcal{N}_w^-$), i.e., if v transmits to w with erasure probability $\varepsilon < 1$. Here we assume that the graph $\mathcal{G}_\varepsilon$ is strongly connected, i.e., that there exists a directed path connecting any pair of its vertices. A distributed computation algorithm on the communication graph $\mathcal{G}_\varepsilon = (\mathcal{V}, \mathcal{E})$ is specified by a pair $\mathcal{A} = (\Phi, \Psi)$ of double-indexed families of maps $\Phi = \{\phi_v^{(t)} : v \in \mathcal{V}, t \in \mathbb{N}\}$, and $\Psi = \{\psi_v^{(t)} : v \in \mathcal{V}, t \in \mathbb{N}\}$, specifically

$$\begin{aligned} \phi_v^{(t)} &: \Theta \times \{0, 1, ?\}^{\mathcal{N}_v^- \times [t-1]} \to \{0, 1\}, \\ \psi_v^{(t)} &: \Theta \times \{0, 1, ?\}^{\mathcal{N}_v^- \times [t]} \to \Theta \, , \end{aligned}$$

and $a_v(t) = \phi_v^{(t)}(i_v(t-1))$, $\hat{y}_v(t) = \psi_v^{(t)}(i_v(t))$.

In the sequel, we shall propose and study some distributed computation algorithms that can be framed in the above general setting. In order to analyze their performance, we will study the distance of the estimates $\hat{y}_v(t)$ from the average of the initial values y:

$$\boldsymbol{e}(t) = \hat{\boldsymbol{y}}(t) - y\mathbf{1} \, .$$

Namely, we define two complexity figures, the *communication complexity* and the *computational complexity*. The communication complexity of a distributed algorithm $\mathcal{A}$ on a graph $\mathcal{G}_\varepsilon$ is measured in terms of the function

$$\tau(\delta) := \inf\left\{t \in \mathbb{N} : \, n^{-1}\mathbb{E}\left[||\boldsymbol{e}(s)||^2\right] \leq \delta, \, \forall s \geq t\right\},$$

where $\delta \in]0, 1]$. In other words, for $\delta \geq 0$, $\tau(\delta)$ denotes the minimum number of binary transmissions each agent has to perform in order to guarantee that the average mean squared estimation error does not exceed δ. Instead, the computational complexity of an algorithm $\mathcal{A}$ on a graph $\mathcal{G}_\varepsilon$ is measured as follows. For every $t \in \mathbb{N}$, and $v \in \mathcal{V}$, we denote by $\kappa_v(t)$ the minimum number of binary operations required by agent v to evaluate the functions $\phi_v^{(t)}(\cdot)$ and $\psi_v^{(t)}(\cdot)$. Then, we define

$$\kappa(\delta) := \max\left\{\sum\nolimits_{t=1}^{\tau(\delta)} \kappa_v(t) \, : \, v \in \mathcal{V}\right\}, \qquad \delta \in]0, 1] \, .$$

Hence, for any $\delta > 0$, $\kappa(\delta)$ denotes the maximum, over all agents $v \in \mathcal{V}$, of the total number of binary operations required to be performed, in order to achieve an average mean squared estimation error not exceeding δ.

6.1 Reliable transmission of continuous information through digital noisy channels

When the communication graph is complete, with all the agents connected through binary erasure broadcast channels, the problem reduces to that of reliable transmission of continuous information through digital noisy channels, which we addressed in Section 2, see [70]. While referring to [70] for general information-theoretical limits and complexity vs performance tradeoffs, for reader's convenience we revise here some results which will be used in the sequel.

Let θ be a random variable taking values in the unitary interval $\Theta = [0, 1]$, according to some a-priori probability law. Consider a memoryless binary erasure channel with erasure probability $\varepsilon \in (0, 1)$. At each time $t \in \mathbb{N}$, the channel has input $a_t \in \{0, 1\}$, output $b_t \in \{0, 1, ?\}$, with b_t conditionally independent from x, $\{a_s, b_s : 1 \leq s \leq t-1\}$, given a_t, and such that $b_t = a_t$ with probability $1 - \varepsilon$, and $b_t =?$ with probability ε. The goal is to design a sequence of encoders $\Upsilon = (\Upsilon_t : \Theta \to \{0, 1\})_{t\in\mathbb{N}}$, and of decoders $\Lambda = (\Lambda_t : \{0, 1, ?\}^t \to \Theta)_{t\in\mathbb{N}}$, such that, if $a_t = \Upsilon_t(x)$, b_t is the corresponding channel output, and $\hat{\theta}_t := \Lambda_t(b_1, \ldots, b_t)$ the current estimate, the mean squared error $\mathbb{E}[(\theta - \hat{\theta}_t)^2]$ is minimized. The computational complexity of the sequential coding scheme (Υ, Λ) is measured, for every time horizon $\ell \in \mathbb{N}$, in terms of the total number k_ℓ of binary operations required to compute $\Upsilon_t(x)$ and $\Lambda_t(b_1, \dots, b_t)$ for all $1 \le t \le \ell$.

Here, in particular, we consider two specific classes of sequential transmission schemes described and analyzed in [70]. The first class is that of random linear tree codes, referred to by the superscript L. These codes have exponential convergence rates with respect to the number of channel uses, and computational complexity proportional to the the cube of the number of channel uses. The second class is that of irregular repetition codes (superscript R). Such codes have linear computational complexity, but subexponential converge rates. The performance of these two classes of codes is summarized in the following lemmas.

Lemma 4 ([70], Coroll. 6.2)**.** *There exist a sequence of linear encoders Υ^L, and a sequence of decoders Λ^L, such that, if $\hat{\theta}_\ell = \Lambda^L_\ell(b_1, \dots, b_\ell)$, then, for all $\ell \ge 0$,*

$$\mathbb{E}\left[(\theta - \hat{\theta}_\ell)^2\right] \le \beta_L^{2\ell}, \qquad k_\ell^L \le B\ell^3, \tag{115}$$

where $\beta_L \in (0,1)$, and $B > 0$ are constants depending on the erasure probability ε only.

Lemma 5 ([70], Prop. 5.1)**.** *There exist a sequence of linear encoders Υ^R, and a sequence of decoders Λ^R, such that, if $\hat{\theta}_\ell = \Lambda^R_\ell(b_1, \dots, b_\ell)$, then, for all $l \ge 0$,*

$$\mathbb{E}\left[(\theta - \hat{\theta}_\ell)^2\right] \le \beta_R^{2\sqrt{\ell}}, \qquad k_\ell^R \le 2\ell, \tag{116}$$

where $\beta_R \in (0,1)$ is a constant depending on the erasure probability ε only.

6.2 Distributed averaging

In this section, we present two iterative distributed averaging algorithms, working on a strongly connected graph $\mathcal{G}_\varepsilon$. Both algorithms are based on a sequence of transmission phases, indexed by $j \ge 1$, alternated to averaging steps. Each agent $v \in \mathcal{V}$ maintains a scalar state $x_v(j)$, $j \ge 0$, which is initialized to the original observation θ_v. The state $x_v(j)$ has to be thought as v's estimate of y at the beginning of the $(j+1)$-th phase. During the j-th transmission phase, each agent broadcasts ℓ_j binary signals to its out-neighbors. These binary signals represent an encoding of the state $x_v(j-1)$. At the end of the j-th phase, each agent estimates each of its in-neighbors' states from the signals received from it, and it updates its state to a convex combination of these estimates and its own current state. The process is then iterated.

We provide now a formal description of the algorithms. Let P be a doubly-stochastic, irreducible matrix adapted to $\mathcal{G}_\varepsilon$, with non-zero diagonal entries. Let $(\ell_j)_{j\in\mathbb{N}}$ be a sequence of positive integers, each ℓ_j representing the length of the j-th transmission phase, and define $h_j := \sum_{i\leq j}\ell_i$, for all $j\in\mathbb{N}$ and $h_0 = 0$. Further, let Υ and Λ be sequences of encoders and decoders as introduced in Sect. 6.1. Then, the proposed distributed algorithms consist of the following steps. First of all, each agent $v\in\mathcal{V}$ initializes its state setting $x_v(0)=\theta_v$. Then, for all $j\in\mathbb{N}$ and $v\in\mathcal{V}$:

Communication phase: v broadcasts an encoded version of its state $x_v(j-1)$ to its out-neighbours, namely, $\forall\, h_{j-1} < t \leq h_j$, it transmits the binary signal

$$a_t = \Upsilon_k\left(x_v(j-1)\right)\,,\qquad k = t - h_{j-1}\,, \tag{117}$$

State update: at the end of the j-th communication phase, v estimates the state of all its in-neighbours, based on the received signals $\{\boldsymbol{b}_v(t)\}_{t=h_{j-1}+1}^{h_j}$; for each $w\in\mathcal{N}_v^-$, let $\hat{x}_w^{(v)}(j-1)$ be the estimate of $x_w(j-1)$ built by agent v, then

$$\hat{x}_w^{(v)}(j-1) = \Lambda_{\ell_j}\left(b_{w\to v}(h_{j-1}+1),\ldots,b_{w\to v}(h_j)\right). \tag{118}$$

Then, v updates its own state according to the following *consensus-like step*:

$$x_v(j) = \sum_{w\in\mathcal{N}_v^-} P_{vw}\hat{x}_w^{(v)}(j-1) + P_{vv}x_v(j-1)\,. \tag{119}$$

Observe that the above-described algorithms can be framed in the general setting described in Sect. 6.1. Indeed, for all $j\geq 1$, one has

$$\begin{aligned} \phi_{h_{j-1}+k}^{(v)}(i_v(h_{j-1}+k)) &= \Upsilon_i\left(x_v(j-1)\right) \qquad & 0<k\leq \ell_j\,,\\ \psi_{h_{j-1}+k}^{(v)}(i_v(h_{j-1}+k)) &= x_v(j-1) \qquad & 0\leq k<\ell_j\,. \end{aligned}$$

Notice that state $x_v(j-1)$ represents the estimate that agent v has of y along all j-th phase, i.e.,

$$\hat{y}_v(t) = x_v(j-1)\,,\qquad \forall h_{j-1}\leq t< h_j\,. \tag{120}$$

In what follows, we consider two implementations of the algorithm. In the first implementation, referred to as algorithm $\mathcal{A}_L$, we use linear tree codes $\Upsilon=\Upsilon^L$, $\Lambda=\Lambda^L$, and phase-lengths $\ell_j^L = S_L j$ for some $S_L\in\mathbb{N}$. In the second implementation, referred to as algorithm $\mathcal{A}_R$, we use repetition codes $\Upsilon=\Upsilon^R$, $\Lambda=\Lambda^R$, and phase-lengths $\ell_j^R = S_R j^2$, for some $S_R \in \mathbb{N}$. Observe that, thanks to (115), one has, for the algorithm $\mathcal{A}_L$,

$$\mathbb{E}\left[\left(\hat{x}_w^{(v)}(j-1) - x_w(j-1)\right)^2\right] \leq \alpha_L^{2j}\,, \tag{121}$$

for every $j \in \mathbb{N}$, $v \in \mathcal{V}$, and $w \in \mathcal{N}_v^-$, where $\alpha_L := \beta_L^{S_L}$. Similarly, for the algorithms $\mathcal{A}_R$, Eq. (116) guarantees that

$$\mathbb{E}\left[\left(\hat{x}_w^{(v)}(j-1) - x_w(j-1)\right)^2\right] \leq \alpha_R^{2j}\,, \tag{122}$$

for every $j \in \mathbb{N}$, $v \in \mathcal{V}$, and $w \in \mathcal{N}_v^-$, where $\alpha_R := \beta_R^{\sqrt{S_R}}$.

It should be mentioned that other choices could have been made for the communication phase lengths, as well as for the coding schemes used during each of them. For instance, block codes of different lengths could have been used during each phase. Our choice of using the same anytime transmission scheme for every agent during each communication phase, has the advantage of fewer memory requirements (only one transmission scheme has to be memorized by each agent), anonymity (each agent uses the same transmission scheme, and the state updating rules only depend on its position in the graph), and adaptiveness with respect to the erasure probability ε. In fact, it is not required to know the actual value of ε in order to design Υ and Λ, see Remarks 3 and 5 in [70].

6.3 Performance analysis

We now present results characterizing the performance of the algorithms $\mathcal{A}_L$, $\mathcal{A}_R$ introduced in Sect. 6.2. Throughout, we assume that $\mathcal{G}_\varepsilon$ is a strongly connected graph, and P is a doubly stochastic, irreducible matrix which is adapted to $\mathcal{G}_\varepsilon$, and has positive diagonal entries. Notice that this implies that P^*P is doubly-stochastic and irreducible. It then follows from Perron-Frobenius theorem that P^*P has the eigenvalue 1 with multiplicity one and corresponding eigenvector $\mathbf{1}$, and all its other eigenvalues have modulus strictly smaller than 1. Hence, P has largest singular value equal to 1 and all other singular values strictly smaller than 1. We denote by $\rho := \rho(P) < 1$ the second largest singular value of P, and assume that $\rho \geq \underline{\rho}$, where $\underline{\rho} > 0$ is some a priori constant.[12]

[12] This may be enforced without using global information, by assuming $P_{vv} \geq (1+\underline{\rho})/2$. Note that this assumption is for analysis' purpose only, and the agents do not need to know $\underline{\rho}$ to run the algorithms. The assumption entails a minimal loss of generality in that it rules out the case $\rho = 0$: related results which cover this case can be found in [64].

Observe that the vector of the estimation errors on y made by the different agents, $\boldsymbol{e}(t) = \hat{\boldsymbol{y}}(t) - y\mathbf{1}$, is constant during each transmission phase, i.e.,

$$\boldsymbol{e}(t) = \boldsymbol{e}(h_j)\,, \qquad \forall\, h_j \leq t < h_{j+1}\,. \tag{123}$$

for any $j \geq 0$. To analyze the performance of our algorithms, it is useful to introduce a suitable decomposition of $\boldsymbol{e}$; for all $j \geq 0$, we can write that

$$\boldsymbol{e}(h_j) = \boldsymbol{z}(j) + \zeta(j)\mathbf{1}\,,$$

where

$$\boldsymbol{z}(j) = \boldsymbol{x}(j) - \left(n^{-1}\mathbf{1}^*\boldsymbol{x}(j)\right)\mathbf{1} \tag{124}$$

represents the difference between the current estimates and the average of the current states, whereas

$$\zeta(j) = n^{-1}\mathbf{1}^*\boldsymbol{x}(j) - y = n^{-1}\mathbf{1}^*\left(\boldsymbol{x}(j) - \boldsymbol{x}(0)\right) \tag{125}$$

accounts for the distance between the current average of the estimates and the average of the initial conditions. Now, observe that the state dynamics (119) may be rewritten in the following compact form

$$\boldsymbol{x}(j+1) = P\boldsymbol{x}(j) + \left(P \odot \Delta(j+1)\right)\mathbf{1}\,, \tag{126}$$

where $\boldsymbol{x}(0) = \boldsymbol{\theta}$ and where $\Delta(j) = (\Delta_{vw}(j))_{v,w\in\mathcal{V}}$ is defined, for all $j \in \mathbb{N}$, by

$$\Delta_{vw}(j) := \begin{cases} \hat{x}_w^{(v)}(j-1) - x_w(j-1) & \text{if} \quad w \in \mathcal{N}_v^- \\ 0 & \text{if} \quad w \notin \mathcal{N}_v^-\,. \end{cases}$$

Notice that, in general, $\Delta_{vw}(j)$ has non-zero mean, and it is not independent from $x_w(j)$, and therefore from the errors introduced by the previous transmission phases $\{\Delta(i)\,:\ 1 \leq i < j\}$. We have the following result.

Proposition 6. *Consider the stochastic system (126), driven by a noise process* $\{\Delta(j)\,:\ j \geq 1\}$ *satisfying*

$$\mathbb{E}[\Delta_{vw}(j)^2] \leq \alpha^{2j}\,, \qquad j \geq 1\,,$$

for some $0 < \alpha < \rho$*. Then, for all* $j \geq 0$*,*

$$\mathbb{E}[\zeta^2(j)] \leq \alpha^2(1-\alpha)^{-2}\,, \tag{127}$$

$$n^{-1}\mathbb{E}[\|\boldsymbol{z}(j)\|^2] \leq \rho^{2j}\left(1 - \alpha/\rho\right)^{-2}\,. \tag{128}$$

The following result characterizes the performance of both algorithms $\mathcal{A}_L$ and $\mathcal{A}_R$.

Theorem 15 (No communication feedback)**.** *For any choice of the initial phase's length S_L (respectively, S_R), there exists a real-valued random variable $\hat{y}$ such that*

$$\mathbb{E}\left[(y-\hat{y})^2\right] \leq \alpha^2(1-\alpha)^{-2}\,, \tag{129}$$

where $\alpha = \beta_L^{S_L}$ (respectively, $\alpha = \beta_R^{\sqrt{S_R}}$) and that the estimates of algorithm $\mathcal{A}_L$ (respectively, $\mathcal{A}_R$) satisfy, with probability one,

$$\lim_{t\to\infty} \hat{y}_v(t) = \hat{y}\,, \qquad \forall v \in \mathcal{V}\,. \tag{130}$$

Moreover, it is possible to choose the initial phase length S_L (respectively, S_R) in such a way that the algorithm $\mathcal{A}_L$ (respectively, $\mathcal{A}_R$) has communication and computational complexities satisfying

$$\tau_L(\delta) \leq C_1 + C_2\frac{\log^3\delta^{-1}}{\log^2\rho^{-1}}\,, \qquad \kappa_L(\delta) \leq C_3 + C_4\frac{\log^7\delta^{-1}}{\log^4\rho^{-1}}\,,$$

and, respectively,

$$\tau_R(\delta) \leq C_5 + C_6\frac{\log^5\delta^{-1}}{\log^3\rho^{-1}}\,, \qquad \kappa_R(\delta) \leq C_7 + C_8\frac{\log^5\delta^{-1}}{\log^3\rho^{-1}}\,,$$

for all $\delta \in]0,1]$, where $\{C_i : i = 1,\dots,8\}$ are positive constants depending on ε only.

Observe that, by (129), the mean squared distance between the asymptotic estimate $\hat{y}$ and the actual value y, is upper bounded by a constant which, quite remarkably, is independent of either the size of the network or the consensus matrix P, and depends only on the length of the first transmission phase. Moreover Theorem 15 shows that both the algorithms $\mathcal{A}_L$ and $\mathcal{A}_R$ have communication and computational complexities growing at most poly-logarithmically in the desired precision. The bounds on the communication (resp. computation) complexities suggest that for the agents it may be sufficient to use fewer channel transmissions in order to achieve a desired precision when running the algorithm $\mathcal{A}_L$ than when running $\mathcal{A}_R$, and that the opposite happens if the number of computations is considered. This behavior has been confirmed in a number of simulations we have run implementing the algorithms. Furthermore, in Theorem 15 both complexities depend on ρ, the second largest singular value of the matrix P. As the matrix P is adapted to the communication graph $\mathcal{G}_\varepsilon$, the dependence of the bounds on ρ captures the effect of the network topology.

7 Conclusion

In this document we reported the advances obtained within the project on the field of control and estimation over noisy communication channels.

In section 2 we analyzed the case of the digital erasure channel and proposed anytime coding algorithms which are suitable to control and estimation purposes. In particular we obtained upper and lower bounds on the highest exponential rate achievable for the mean squared error with respect to the number of channel uses. Using finer information-theoretic arguments, most of our results can be extended to more general discrete memoryless channels. In particular, Theorem 1 can be extended to general discrete memoryless channels. Two coding methods have been proposed. One is very computationally demanding, but which optimally exploits the communication resource (exponential error rates). The other is instead suboptimal from the point of view of the use of communication resource (subexponential error rates), but is on the other hand extremely algorithmically simple. Some of the questions raised in this section have been left open. A particularly relevant issue is to design other possible encoding methods with intermediate performance compared with the previous ones, and in particular the design, if they exist, of low-complexity coding schemes achieving exponential error rates.

In section 5 we have investigated joint optimization of the encoder and the controller in closed-loop control of a linear plant with low-rate feedback over a memoryless noisy channel. We introduced an iterative approach to the design of encoder–controller pairs inspired by the traditional design of vector quantizers. In the case of full SI, we introduced a "virtual help-system," the open-loop encoder system. We showed that a CE controller is optimal for any given encoder in this system, and we argued that encoder–controller pairs designed for the help-system can be translated to perform well in the original system. In the case of partial SI, we cannot claim that enforcing the CE controller structure is without loss. However, since the general controller problem is challenging in this case, we used CE controllers as sub-optimal, but practically feasible approximations. Finally, we have performed various numerical investigations. Our results demonstrate the promising performance obtained by employing the proposed design algorithms.

In section 3, we studied the problem of mean square stabilizing a discrete time LTI system over some basic topologies of Gaussian sensor networks. We proposed to use delay-free linear communication and control strategies, and thereby obtained sufficient conditions for stabilization. We also obtained necessary conditions for stabilization using information theoretic bounds and in some cases bounds are shown to be tight. Our results reveal a relationship between the communication channel parameters (i.e., signal-to-noise ratios) and the possibility of stabilizing the plant. Some discussion on vector valued systems can be found in [41]. An interesting extension of this work would be to consider instantaneous non-linear relaying strategies which can potentially increase the achievable rate and thus extend the class of stablizable systems over the considered Gaussian sensor networks.

In section 4 we studied the problem of mean square stabilizing two discrete time scalar LTI systems in closed-loop via control over white Gaussian multiple-access, broadcast, and interference communication channels. We propose to use simple linear communication and control schemes which whiten the state process and make it Gaussian, and therefore the optimal decoding of the transmitted state values at the remote control unit(s) is linear and memoryless. The stability regions obtained are associated with the achievable rate regions for the given channels with noiseless feedback. Therefore our results reveal relationship between mean square stability of the two plants and the communication channels' parameters, i.e., average power consumed by the encoder(s) and the average power of the noise components in different links. The stability results provided can be easily extended for the setup where the links from the controller(s) to the plants are also white Gaussian communication channels. For this setup we can have an encoder at each control unit to encode the control action and an MMSE decoder at each plant to decode the transmitted value of the control action. As long as the encoders, the decoders, and the controllers are linear, the nature of the problem does not change and the stability results can be easily obtained [15].

Finally in section 6 we have considered the averaging problem on networks of digital links, and established suitable performance figures to evaluate its algorithmic solutions, in terms of communication and computation complexities. On this ground, our main contribution has consisted in proposing and analyzing a family of average consensus algorithms, based on encoding/decoding schemes with precision increasing with time. Such increase is meant to compensate the effect of errors in digital communications, which can be modeled as additive noise. Our results show almost sure convergence to average consensus, with communication and computation complexities growing poly-logarithmically in the desired precision. The question is open whether a logarithmic algorithm can be designed for average consensus on digital networks, and how much global information it would require to be run by the agents. Moreover we plan also to combine the technique presented here with the ZIZO algorithm introduced above to deal with the noiseless communication channel.

References

[1] A. A. ZAIDI, T. J. OECHTERING AND M. SKOGLUND, *Rate Sufficient Conditions for Closed–loop Control over AWGN Relay Channels*, IEEE ICCA (2010).

[2] G. KRAMER, *Feedback strategies for white Gaussian interference networks*, IEEE Trans. Inform. Theory, 48 (2002), pp.1423–1438.

[3] S. YÜKSEL, *Stochastic Stabilization of Noisy Linear Systems with Fixed-Rate Limited Feedback*, IEEE Trans. Automat. Control, 99 (2010).

[4] H. SATO, *The capacity of the Gaussian interference channel under strong interference*, IEEE Trans. Inform. Theory, 27 (1981), pp.786–788.

[5] M. GASTPAR AND A. LAPIDOTH AND M. WIGGER, *When Feedback doubles the prelog in AWGN networks*, IEEE Trans. Inform. Theory, *Submitted*, available at *http://arxiv.org/abs/1003.6082*.

[6] A. CARLEIAL, *Interference Channels*, IEEE Trans. Inform. Theory, 24 (1978), pp.60–70.

[7] A. A. ZAIDI, S. YÜKSEL, T. J. OECHTERING AND M. SKOGLUND, *On Optimal Policies for Control and Estimation Over a Gaussian Relay Channel*, IEEE CDC (2011).

[8] A. A. ZAIDI, T. J. OECHTERING AND M. SKOGLUND, *Closed-loop stabilization Over Gaussian Interference Channel*, IFAC World Congress (2011).

[9] A. A. ZAIDI, T. J. OECHTERING, S. YÜKSEL AND M. SKOGLUND, *Sufficient conditions for closed-loop control over a Gaussian relay channel*, IEEE ACC (2011).

[10] A. A. ZAIDI, T. J. OECHTERING, M. SKOGLUND AND S. YÜKSEL, *Closed-loop Control over Half-duplex AWGN Relay Channels*, Reglermöte (2010).

[11] A. A. ZAIDI, T. J. OECHTERING AND M. SKOGLUND, *Rate Sufficient Conditions for Closed–loop Control over AWGN Relay Channels*, IEEE ICCA (2010), pp.602–607.

[12] A. A. ZAIDI, T. J. OECHTERING AND M. SKOGLUND, *Sufficient Conditions for Closed-Loop Control Over Multiple-Access and Broadcast Channels*, IEEE CDC (2010).

[13] S. YÜKSEL AND S. TATIKONDA, *Distributed Sensing and Control over Communication Channels and a Counterexample*, Allerton Conference on Communication, Control, and Computing (2007), pp.260–267.

[14] S. YÜKSEL AND T. LINDER, *Optimization and convergence of observation channels in stochastic control*, SIAM J. Control Optim, *In press*.

[15] S. YÜKSEL AND T. BAŞAR, *Achievable Rates for Stability of LTI Systems over Noisy Forward and Feedback Channels*, Proc. Conference on Information Sciences and Systems (2005).

[16] S. YÜKSEL AND T. BAŞAR, *Control over Noisy Forward and Reverse Channels*, IEEE Trans. Automat. Control, 56 (2011).

[17] S. YÜKSEL, *On Optimal Causal Coding of Partially Observed Markov Sources under Classical and Non-Classical Information Structures*, IEEE ISIT, 2010, pp.81–85.

[18] Y. WU AND S. VERDÚ, *Functional properties of MMSE*, IEEE ISIT, 2010, pp.1453–1457.

[19] H. S. WITSENHAUSEN, *A counterexample in stochastic optimum control*, SIAM J. Control, 6 (1968), pp.131–147.

[20] J. WALRAND AND P. VARAIYA, *Optimal causal coding-decoding problems*, IEEE Trans. Inform. Theory, 29 (1983), pp.814–820.

[21] R. RADNER, *Team Decision Problems*, Annals of Mathematical Statistics, 33 (1962), pp.857–881.

[22] L. OZAROW, *The capacity of the white Gaussian multiple-access channel with feed-back*, IEEE Trans. Inform. Theory, 30 (1984), pp.623–629.

[23] L. OZAROW AND S. LEUNG-YAN-CHEONG, *An achievable region and outer bound for the Gaussian broadcast channel with feedback*, IEEE Trans. Inform. Theory, 30 (1984), pp.667–671.

[24] P. MINERO, M. FRANCESCHETTI AND S. DEY AND G. NAIR, *Data rate theorem for stabilization over time varying feedback channels*, IEEE Trans. Automat. Control, 54 (2009), pp.243–255.

[25] A. MATVEEV AND A. SAVKIN, *Comments on Control over Noisy Channels and Relevant Negative Results*, IEEE Trans. on Automat. Control, 50 (2005), pp. 2105–2110.

[26] J. L. MASSEY, *Causality, Feedback and Directed Information*, IEEE ISIT, 1990.

[27] G. M. LIPSA AND N. C. MARTINS, *Finite horizon optimal memoryless control of a delay in Gaussian noise: A simple counterexample.*, IEEE CDC (2008), pp.1628–1635.

[28] G. M. LIPSA AND N. C. MARTINS, *Optimal memoryless control in Gaussian noise: A simple counterexample*, Automatica (2011), *In press*.

[29] U. KUMAR AND V. GUPTA AND J. N. LANEMAN, *Sufficient conditions for stabilizability over Gaussian relay channel and cascade channels*, IEEE CDC (2010), pp.4765–4770.

[30] K. S. GOMADAM, AND S. A. JAFAR, *Optimal relay functionality for SNR maximization in memoryless relay networks*, IEEE Journal on Selected Areas in Communications, 25 (2007), pp.390–401.

[31] G. KRAMAER, M. GASTPAR, AND P. GUPTA, *Cooperative strategies and capacity theorems for relay networks*, IEEE Trans. Inform. Theory, 51 (2005), pp.3037–3063.

[32] J. KARLSSON AND M. SKOGLUND, *Optimized low-delay source-channel-relay mapping*, IEEE Trans. Commun., 58 (2010), pp.1397–1404.

[33] Y. C. HO, *Team decision theory and information structures*, Proceedings of the IEEE, 68 (1980), pp.644–654.

[34] M. HAYES, *Statistical digital signal processing and modelling*, John Wiley Sons, Inc., (1996).

[35] M. GASTPAR AND M. VETTERLI, *On the capacity of large Gaussian relay networks*, IEEE Trans. Inf. Theory, 51 (2005), pp.765–779.

[36] M. GASTPAR, B. RIMOLDI AND M. VETTERLI, *To code or not to code*, IEEE ISIT, 2000.

[37] M. GASTPAR, *Cut-set arguments for source-channel networks*, IEEE ISIT, 2004.

[38] M. GASTPAR, *Uncoded Transmission Is Exactly Optimal for a Simple Gaussian Sensor Network*, IEEE Trans. Inf. Theory, 54 (2008), pp.5247–5251.

[39] DENIZ GÜNDÜZ, CHRIS T. K. NG, ELZA ERKIP AND ANDREA J. GOLDSMITH, *Source Transmission over Relay Channel with Correlated Relay Side Information*, IEEE ISIT, (2007).

[40] J. H. BRASLAVSKY, R. H. MIDDLETON AND J. S. FREUDENBERG, *Feedback Stabilization over Signal-to-Noise Ratio Constrained Channels*, IEEE Trans. Automat. Control, 52 (2007), pp.1391–1403.

[41] A. A. ZAIDI, T. J. OECHTERING, S. YÜKSEL AND M. SKOGLUND, *Linear Stabilization Over Gaussian Sensor Networks*, IEEE Trans. Automat. Control, *In preparation*.

[42] J. FREUDENBERG, R. MIDDLETON, AND V. SOLO, *Stabilization and Disturbance Attenuation Over a Gaussian Communication Channel*, IEEE Trans. Automat. Control, 55 (2010), pp. 795–799.

[43] N. ELIA, *Optimum mean-square quantization of a noisy input*, IEEE Trans. Inf. Theory, 11 (1965), pp. 293–294.

[44] N. ELIA, *When Bode meets Shannon: control–oriented feedback communication schemes*, IEEE Trans. on Automat. Control, 49 (2004), pp. 1477–1488.

[45] R. L. DOBRUSHIN AND B. S. TSYBAKOV, *Information Transmission with additional noise*, IRE Trans. Inf. Theory, 18 (1962), pp. 293–304.

[46] T. COVER AND J. THOMAS, *Elements of information theory*, John Wiley Sons, Inc., (2006).

[47] TOBY BERGER, *Rate Distortion Theory: A mathematical basis for data compression*, Englewood Cliffs, NJ: Prentice-Hall, (1971).

[48] R. Bansal and T. Başar, *Solutions to a class of linear-quadratic-Gaussian LQG stochastic team problems with nonclassical information*, Systems Control Letters, 9 (1987), pp. 125–130.

[49] E. Ayanoglu and R. M. Gray, *The design of joint source and channel trellis waveform coders*, IEEE Transactions on Information Theory, 33 (1987), pp. 855-865.

[50] M. Aoki, *Optimization of Stochastic Systems - Topics in Discrete-Time Systems*, Academic Press, 1967.

[51] T. C. Aysal, M. Coates, and M. Rabbat, *Distributed average consensus using probabilistic quantization*, in IEEE Workshop on Statistical Signal Processing, Maddison, Wisconsin, Aug. 2007, pp. 640–644.

[52] G. Baldan and S. Zampieri, *An efficient quantization algorithm for solving average-consensus problems*, in Proceedings of the European Control Conference, 2009.

[53] R. Bansal and T. Basar, *Simultaneous design of measurement and control strategies for stochastic systems with feedback*, Automatica, 25 (1989), pp. 679–694.

[54] L. Bao, M. Skoglund, and K. H. Johansson, *Iterative encoder–controller design for feedback control over noisy channels*, (2010).

[55] E.R. Berlekamp, R.J. McEliece, and H.C.A. Van Tilborg, *On the inherent intractability of certain coding problems*, IEEE Trans. Inform. Theory, 24(1978), pp. 384–386.

[56] D. Bertsekas, and J. Tsitsiklis, Parallel and Distributed Computation: Numerical Methods, Athena Scientific, Belmont, MA, 1997.

[57] B. Bollobas, Modern graph theory, Springer Verlag, New York, NY, 1998.

[58] S. Borade, B. Nakiboğlu, and L. Zheng, *Unequal error protection: some fundamental limits*, IEEE Trans. Inform. Theory, 55 (2009), pp. 5511–5539.

[59] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, *Randomized gossip algorithms*, IEEE Transactions on Information Theory, 52 (2006), pp. 2508–2530.

[60] J. H. BRASLAVSKY, R. H. MIDDLETON, AND J. S. FREUDENBERG, *Feedback stabilization over signal-to-noise ratio constrained channels*, 52 (2007), pp. 1391–1403.

[61] R.-W. BROCKETT AND D. LIBERZON, *Quantized feedback stabilization of linear systems*, IEEE Transactions on Automatic Control, 45 (2000), pp. 1279–1289.

[62] S. BROSS AND M. WIGGER, *On the relay channel with receiver-transmitter feedback*, IEEE Trans. Inform. Theory, 55 (2009), pp. 275–291.

[63] M.V. BURNASHEV, *A new lower bound for the α-mean error of parameter transmission over the white Gaussian channel*, IEEE Trans. Inform. Theory, 30(1984), pp. 23–34.

[64] R. CARLI, G. COMO, P. FRASCA, AND F. GARIN, *Average consensus over digital noisy networks*, Proc. IFAC NecSys'09, Venezia (Italy), 24–26 Sept. 2009, pp. 36–41.

[65] ——, *Distributed averaging on digital erasure channels*, Submitted to Automatica, provisionally accepted.

[66] R. CARLI, F. FAGNANI, P. FRASCA, T. TAYLOR, AND S. ZAMPIERI, *Average consensus on networks with transmission noise or quantization*, in ECC '07, Kos, 2007, pp. 1852–1857.

[67] R. CARLI, P. FRASCA, F. FAGNANI, AND S. ZAMPIERI, *Gossip consensus algorithms via quantized communication*, Automatica, 46 (2010), pp. 70–80.

[68] C. D. CHARALAMBOUS AND A. FARHADI, *LQG optimality and separation principle for general discrete time partially observed stochastic systems over finite capacity communication channels*, Automatica, 44 (2008), pp. 3181–3188.

[69] G. COMO, AND M. DAHLEH, *Lower bounds on the estimation error in problems of distributed computation*, invited paper, in Proc. of ITA Workshop, (San Diego, CA, US), Feb. 8–13, 2009, pp. 70–76.

[70] G. COMO, F. FAGNANI, AND S.ZAMPIERI., *Anytime reliable transmission of real-valued information through digital noisy channels*, SIAM Journal on Control and Optimization, Vol. 48 (2010), pp. 3903-3924.

[71] T. M. COVER, AND J. A. THOMAS, Elements of Information Theory, John Wiley and Sons, New York, NY, 1991.

[72] I. CSISZÀR, *Joint source-channel error exponent*, Probl. Control Inf. Theory, 9 (1980), pp. 315–328.

[73] ——, *On the error exponent of source-channel transmission with a distortion threshold*, IEEE Trans. Inform. Theory, 28 (1982), pp. 823–828.

[74] G. CYBENKO, *Dynamic load balancing for distributed memory multiprocessors*, Journal of parallel and distributed computing, 7 (1989), pp. 279–301.

[75] R. DIEKMANN, A. FROMMER, AND B. MONIEN, *Efficient schemes for nearest neighbor load balancing*, Parallel computing, 25 (1999), pp. 789–812.

[76] A.G. DIMAKIS, A.D. SARWATE, AND M.J. WAINWRIGHT, *Geographic Gossip: Efficient Averaging for Sensor Networks*, IEEE Trans. Sig. Proc., 56 (2008), pp. 1205–1216.

[77] F. FAGNANI AND S. ZAMPIERI, *Randomized consensus algorithms over large scale networks*, IEEE J. on Selected Areas of Communications, 26 (2008), pp. 634–649.

[78] N. FARVARDIN, *A study of vector quantization for noisy channels*, IEEE Trans. on Information Theory, 36 (1990), pp. 799–809.

[79] G.D. FORNEY JR., *Convolutional codes II. Maximum-likelihood decoding*, Inf. Control, 25 (1974), pp. 222-266.

[80] P. FRASCA, R. CARLI, F. FAGNANI, AND S. ZAMPIERI, *Average consensus by gossip algorithms with quantized communication*, in Proc. of CDC, Cancun, Mexico, 2008, pp. 4831–4836.

[81] ——, *Average consensus on networks with quantized communication*, International Journal of Robust and Nonlinear Control, 19 (2009), pp. 1787-1816.

[82] J. FREUDENBERG, R. MIDDLETON, AND V. SOLO, *The minimal signal-to-noise ratio rqeuired to stabilize over a noisy channel*, in Proc. ACC, 2006.

[83] M. FU, *Linear quadratic gaussian control with quantized feedback*, in Proc. of the 2009 American Control conference, June 2009.

[84] R. G. GALLAGER, Information Theory and Reliable Communication, Wiley, New York, NY, 1968.

[85] R. G. GALLAGER AND B. NAKIBOGLU, *Variations on a theme by Schalkwijk and Kailath*, IEEE Trans. Inf. Theory, 56 (2010).

[86] A. GERSHO AND R. M. GRAY, *Vector quantization and signal compression*, Kluwer, 1992.

[87] M. HIRSCH AND S. SMALE, *Differential Equations, Dynamical Systems, and Linear Algebra*, San Diego, CA: Academic, 1974.

[88] J.-P. HESPANHA , A. ORTEGA , AND L. VASUDEVAN , *Towards the control of linear systems with minimum bit–rate*, in 15th Int. Symp. Mathematical Theory of Networks and Systems (MTNS), Notre Dame, IL, USA, 2002.

[89] C. INTANAGONWIWAT, R. GOVINDAN, AND D. ESTRIN, *Directed diffusion: a scalable and robust communication paradigm for sensor networks*, In Proc. ACM/IEEE Conf. Mobile Computing and Networking, 2000, pp. 56–67.

[90] A. JADBABAIE, J. LIN, AND A.S. MORSE, *Coordination of groups of mobile autonomous agents using nearest neighbor rules*, IEEE Trans. Automat. Control, 48 (2003), pp. 988–1001.

[91] J. JAGLIN, C. CANUDAS-DE-WIT, AND A. FARHADI, *Dwell - time adaptive delta modulation signal coding for networked controlled systems*, to be submitted shortly to the IEEE Transactions on Automatic Control.

[92] A. KASHYAP, T. BAŞAR, AND R. SRIKANT, *Quantized consensus*, Automatica, 43 (2007), pp. 1192–1203.

[93] D. KEMPE, A. DOBRA, AND J. GEHRKE, *Gossip-based computation of aggregate information*, Proc. IEEE FOCS 2003, pp. 1–10.

[94] G. KRAMAER, I. MARIC, AND R. YATES, *Cooperative communications*, Foundations and Trends® in Networking, 1 (2006), pp. 271 –425.

[95] J. LAVAEI AND R. M. MURRAY, *On quantized consensus by means of gossip algorithm – Part I: Convergence proof*, in Proc. of the American Control Conference, St. Louis, USA, July 2009.

[96] ——, *On quantized consensus by means of gossip algorithm – Part II: Convergence time*, in Proc. of the American Control Conference, St. Louis, USA, 2009.

[97] D. LIBERZON, *On stabilization of linear systems with limited information*, IEEE Transactions on Automatic Control, 48 (2003), pp. 304–307.

[98] D. LIBERZON AND D. NESIC, *Input to state stabilization of linear systems with quantized state measurements*, IEEE Transactions on Automatic Control, Volume 52 no. 5 (2007), pp. 767–781.

[99] M. LUBY, *LT codes*, in Proc. 43rd IEEE FOCS 2002, November 16-19, 2002, pp. 271–282.

[100] B. MASNICK, AND J. WOLF, *On linear unequal error protection*, IEEE Trans. Inf. Theory, 13 (1967), pp. 600–607.

[101] A.S. MATVEEV, AND A.V. SAVKIN, *Estimation and control over communication networks*, Birkhauser, Boston, MA, 2009.

[102] A. MATVEEV AND A. SAVKIN, *An analogue of Shannon information theory for detection and stabilization via noisy discrete communication channels*, SIAM J. Control Optim., 46 (2007), pp. 1323–1367.

[103] A. S. MATVEEV AND A. V. SAVKIN, *The problem of LQG optimal control via a limited capacity communication channel*, Systems and Control letters, 53 (2004), pp. 51–64.

[104] D. MCKAY, Information theory, inference, and learning algorithms, Cambridge University Press, Cambridge, UK, 2003.

[105] R. MIDDLETON, A. ROJAS, J. FREUDENBERG, AND J. BRASLAVSKY, *Feedback stabilization over a first order moving average gaussian noise channel*, IEEE Trans. Automat. Control, 54 (2009).

[106] L. MOREAU, *Stability of multiagent systems with time-dependent communication links*, IEEE Trans. Automat. Control, 50 (2005), pp. 169–182.

[107] S. MUTHUKRISHNAN, B. GHOSH, AND M. SCHULTZ, *First and second order diffusive methods for rapid, coarse, distributed load balancing*, Theory of computing systems, 31 (1998), pp. 331–354.

[108] G. N. NAIR AND R. J. EVANS, *Mean square stabilisability of stochastic linear systems with data rate constraints*, in Proc. IEEE CDC, 2002.

[109] ——, *Stabilizability of stochastic linear systems with finite feedback data rates*, SIAM J. Control Optim., 43 (2004), pp. 413–436.

[110] G. NAIR, F. FAGNANI, S. ZAMPIERI, AND R. EVANS, *Feedback control under data rate constrints: An overview*, Proceeding of the IEEE, 95 (2007), pp. 108–136.

[111] R. OLFATI-SABER, AND R.M. MURRAY, *Consensus problems in networks of agents with switching topology and time-delays*, IEEE Trans. Automat. Control, 49 (2004), pp. 1520–1533.

[112] M. PENROSE, *Random Geometric Graphs*, Oxford Studies in Probability, Oxford University Press, 2003.

[113] W. REN, AND R.W. BEARD, *Consensus seeking in multiagent systems under dynamically changing interaction topologies*, IEEE Trans. Automat. Control, 50 (2005), pp. 655–661.

[114] T.J. RICHARDSON, AND R. URBANKE, *Modern Coding Theory*, Cambridge Univeristy Press, Cambridge, UK, 2007.

[115] A. SAHAI, *Why do block length and delay behave differently if feedback is present*?, IEEE Trans. Inform. Theory, 54 (2008), pp. 1860–1886.

[116] A. SAHAI AND S. MITTER, *The necessity and sufficiency of anytime capacity for stabilization of a linear system over noisy communication links–Part I: Scalar systems*, IEEE Trans. Inform. Theory, 52 (2006), pp. 3369–3395.

[117] ——, *Source coding and channel requirements for unstable processes*, IEEE Trans. Inform. Theory, (2006), submitted.

[118] SCHALKWIJK AND T. KAILATH, *A coding scheme for additive noise channels with feedback–i: No bandwidth constraint*, IEEE Trans. Inform. Theory, 12 (1966), pp. 172–182.

[119] Y. SHARON AND D. LIBERZON, *Input to state stabilization with number of quantization regions*, in Conference on Decision and Control, 2007.

[120] E. SILVA, M. DERPICH, J. OSTERGAARD, AND D. QUEVEDO, *Simple coding for achieving mean square stability over bit-rate limited channels*, in Proc. IEEE CDC, 2008.

[121] T. SIMSEK, R. JAIN, AND P. VARAIYA, *Scalar estimation and control with noisy binary observation*, IEEE Trans. Automat. Control, 49 (2004), pp. 1598–1603.

[122] A. TAHBAZ-SALEHI AND A. JADBABAIE, *A necessary and sufficient condition for consensus over random networks*, IEEE Transactions on Automatic Control, 53 (2008), pp. 791–795.

[123] S. TATIKONDA AND S. MITTER, *Control over noisy channels*, IEEE Trans. Automat. Control, 49 (2004), pp. 1196–1201.

[124] ——, *Control under communication constraints*, IEEE Trans. Automat. Control, 49 (2004), pp. 1056–1068.

[125] L.N. TREFETHEN, AND D. BAU, III, Numerical linear algebra, SIAM, Philadelphia, PA, 1997.

[126] J.N. TSITSIKLIS, Problems in decentralized decision making and computation, Ph.D. dissertation, Dep. Elec. Eng. Comput. Sci., Mass. Inst. Technol., Cambridge, MA, 1984.

[127] A.J. VITERBI, Further results on optimal decoding of convolutional codes, *IEEE Trans. Inform. Theory*, 15 (1969), pp. 732–734.

[128] J.-J. XIAO AND Z.-Q. LOU, *Decentralized estimation in an inhomogeneous sensing environment*, IEEE Trans. Inf. Theory, 51 (2005), pp. 3564–3575.

[129] S. YUKSEL AND T. BASAR, *Coding and control over discrete noisy forward and feedback channels*, in Proc. IEEE CDC, 2005.

[130] S. YÜKSEL AND S. TATIKONDA, *A counterexample in distributed optimal sensing and control*, IEEE Trans. Automat. Control, 54 (2009).

[131] A. A. ZAIDI, M. N. KHORMUJI, S. YAO, AND M. SKOGLUND, *Rate-maximizing mappings for memoryless relaying*, in Proc. IEEE ISIT, 2009.

[132] A. A. ZAIDI, T. J. OECHTERING, AND M. SKOGLUND, *Rate sufficient conditions for closed–loop control over AWGN relay channels*, in IEEE ICCA, 2010.

[133] A. A. ZAIDI, T. J. OECHTERING, M. SKOGLUND, AND S. YÜKSEL, *Rate sufficient conditions for closed–loop control over half–duplex AWGN relay channels*, in Reglermöte, 2010.

[134] J. ZHAO, R. GOVINDAN, AND D. ESTRIN, *Computing aggregates for monitoring wireless sensor networks*, Proc. SNPA 2003, pp. 139–148.

[135] S. ZILBERSTEIN, *Using Anytime Algorithms in Intelligent Systems*, AI Magazine, 17 (1996), pp. 73–83.